

(19) 世界知的所有権機関
国際事務局(43) 国際公開日
2000年12月14日 (14.12.2000)

PCT

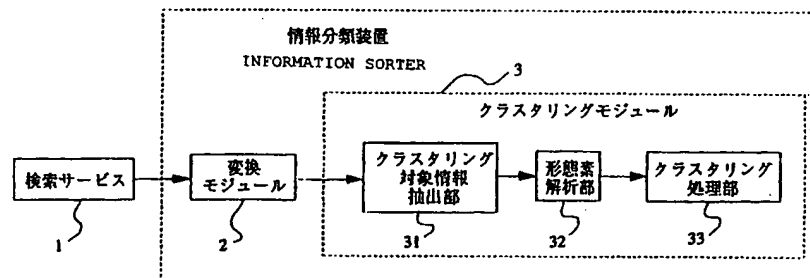
(10) 国際公開番号
WO 00/75809 A1

- (51) 国際特許分類: G06F 17/30
- (21) 国際出願番号: PCT/JP00/03623
- (22) 国際出願日: 2000年6月2日 (02.06.2000)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ: 04 Feb 01
特願平11/158497 1999年6月4日 (04.06.1999) JP
特願平11/171723 1999年6月17日 (17.06.1999) JP
特願平11/193141 1999年7月7日 (07.07.1999) JP
- (71) 出願人 (米国を除く全ての指定国について): セイコーエプソン株式会社 (SEIKO EPSON CORPORATION) [JP/JP]; 〒163-0811 東京都新宿区西新宿2丁目4番1号 Tokyo (JP).
- (72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 長石道博 (NAGAISHI, Michihiro) [JP/JP], 三輪真司 (MIWA, Shinji) [JP/JP]; 〒392-8502 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内 Nagano (JP).
- (74) 代理人: 鈴木喜三郎, 外 (SUZUKI, Kisaburo et al.); 〒392-8502 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社 知的財産部内 Nagano (JP).
- (81) 指定国 (国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) 指定国 (広域): ARIPO 特許 (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), ユーラシア特許 (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ特許

[続葉有]

(54) Title: INFORMATION SORTING METHOD, INFORMATION SORTER, RECORDED MEDIUM ON WHICH INFORMATION SORTING PROGRAM IS RECORDED

(54) 発明の名称: 情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体



- 1...SEARCH SERVICE
2...CONVERSION MODULE
3...CLUSTERING MODULE
31...CLUSTERING OBJECT INFORMATION EXTRACTING SECTION
32...MORPHEME ANALYZING SECTION
33...CLUSTERING SECTION

(57) Abstract: There has been conventionally a problem that a large amount of information obtained by search using a key word inputted by a user through a multipurpose search service is only presented in a list and it is difficult to grasp the contents of the results of search if the amount of results of search

[続葉有]

WO 00/75809 A1



(AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI 特許 (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

添付公開書類:

— 国際調査報告書

is large. A clustering module (3) obtains the results of search by a multipurpose search service (1), clusters the search results, and presents the results of the clustering. A converting module (2) is provided for converting the results of the search to a format which the clustering module can process.

(57) 要約:

汎用の検索サービスを用いて検索を行う場合、ユーザの入力するキーワードに対して検索された多数の情報は羅列的に提示されるだけであり、検索結果が多いと検索結果の内容の把握がしにくいという課題があった。

汎用の検索サービス1で検索された複数の検索結果をクラスタリングモジュール3が取得して、そのクラスタリングモジュール3では、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を提示する。このとき、検索サービス1で検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュール2を備える。

明細書

情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体

技術分野

本発明は汎用の検索サービスで検索された結果に対しクラスタリング処理を施すことで、ユーザに見易い形で提示するようにした情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体に関する。

背景技術

ネットワーク上に存在する膨大な量の情報の中からユーザの所望とする情報を検索する場合、検索サービスの存在は重要である。たとえば、インターネットで web ページを検索する際、ユーザは、幾つかの検索サービスの中から任意の検索サービスを選び、自分の欲しい情報を得るための検索要求としてのキーワードを入力する。これによって、検索サービス側では、入力されたキーワードに基づいて情報検索を行って、その検索結果をユーザに提示する。

しかし、検索サービスによって検索される情報は膨大な量となることも多く、その中からユーザの本当に欲しい情報を見つけるのは非常に大変である。近年、web ページは増大の一途をたどっているため、検索された多数の情報を如何にユーザにわかりやすく提示するかが大きな課題となっている。

最近では、検索された情報をユーザが見やすい形に加工して提示する手法も開発され実用化されつつある。たとえば、ユーザの入力したキーワードで検索された結果から得られるキーワードを用いて再検索することで、情報の絞り込みを行い、ユーザの所望とする web ページを見つけやすくする方法がある。つまり、検索によって得られる検索結果の集合を特徴づけるキーワードを抽出して、ユーザの本当に欲しい情報の集合に収束させる方法である。

このように、膨大な情報の中から、似た情報の集合を見つけることをクラスタリングという。情報処理ではこのクラスタリングはよく知られた手法であり、膨

大な文書を分類する場合などに一般に使われている。

しかし、現在、一般のユーザに広く利用されている検索サービス（汎用検索サービスという）が検索した結果をクラスタリングすることは行われてはなく、前述したように、入力されたキーワードに基づいて情報を抽出して、抽出された情報を羅列的にユーザに提示するというのが一般的である。したがって、ユーザは、その羅列された多数の情報から自分の欲しい情報を探すという面倒な作業をせざるを得なかった。

そこで本発明は、汎用の検索サービスで得られた検索結果をクラスタリングすることで、ユーザに対し見易い形での検索情報提示を可能とすることを目的としている。

図面の簡単な説明

図 1 は、本発明の第 1 の実施形態を説明する図であり、ある 1 つの検索サービスで得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

図 2 は、本発明の第 1 の実施形態で用いられる、ある検索サービスで検索された検索結果としての複数の文書例を示す図である。

図 3 は、図 1 で示されたクラスタリング処理部の構成を示すブロック図である。

図 4 は、第 1 の実施形態における文書分類処の手順を概略的に説明するフローチャートである。

図 5 は、図 2 で示されたそれぞれの文書のタイトルから抽出された特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルの内容の示す図である。

図 6 は、図 5 の特徴テーブルを基にしてそれぞれの文書を分類した分類結果を示す図である。

図 7 は、図 6 の分類結果に基づいてそれぞれの文書のタイトルをクラスタリングした例を示す図である。

図 8 は、選択されたある 1 つの検索サービスから得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

図 9 は、複数の検索サービスから得られた検索結果に対してクラスタリングを可能とする場合の情報分類装置の構成を示すものである。

図 10 は、本発明の第 2 の実施形態を説明する構成図である。

図 11 は、ある検索サービスで検索された検索結果としての複数の文書をクラスタリングした結果の一例を示す図である。

図 12 は、本発明の第 2 の実施形態における情報分類処理手順を概略的に説明するフローチャートである。

図 13 は、図 11 で示されたクラスタリング結果をクラスタ順位再構成処理した結果を示す図である。

図 14 は、本発明の第 3 の実施の形態を説明する構成図である。

図 15 は、本発明の第 3 の実施形態における情報分類処理手順を概略的に説明するフローチャートである。

図 16 は、図 11 で示されたクラスタリング結果とその概要一覧表を表示した例を示す図である。

図 17 は、URL アドレスでクラスタリング処理して得られたクラスタリング結果とその概要一覧表を表示した例を示す図である。

発明の開示

前述の目的を達成するために、本発明の情報分類方法は、検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力するようにしている。

この情報分類方法において、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えるようにしている。

そして、前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられる。

このとき、複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行

うことも可能であり、また、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことも可能であり、さらに、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことも可能である。

また、前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとしている。

また、本発明の情報分類方法において、前記クラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理結果のクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力することもできる。

さらに、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの平均値をそれぞれのクラスタごとに求め、クラスタごとの平均値をそれぞれのクラスタのクラススコアとし、そのクラススコアによって、クラスタの順位を再構成するものである。

また、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの最大値をそれぞれのクラスタごとに得て、そのクラスタごとのスコアの最大値をそれぞれのクラススコアとし、そのクラススコアによって、クラスタの順位を再構成してもよい。

また、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書がそれぞれの文書対応に付されたスコアの大きい順に並べられている場合、その中央または中央付近に位置するスコアをそれぞれのクラスタごとに得て、そのクラスタごとの中央または中央付近に位置するスコアをそれぞれのクラススコアとし、そのクラススコアによって、クラスタの順位を再構成してもよい。

また、前記クラスタリング処理を複数の検索サービスによって得られた検索結果に対応して行うことを可能とする場合、前記クラスタの順位を再構成するため

のクラスタスコアを求める処理は、複数の検索サービスそれぞれに対応して行ってもよい。

また、前記クラスタリング処理は、それぞれの文書のタイトルを検出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいて行ってもよい。

また、前記クラスタ順位が再構成されたクラスタリング結果の出力の仕方は、クラスタスコアの高いクラスタ順に表示し、クラスタスコアが同じであるクラスタが存在する場合には、クラスタ内の文書数の多いクラスタを高順位とすることができる。

また、本発明の情報分類方法において、前記クラスタリング処理結果に対し、クラスタリング処理結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング処理結果とともに出力してもよい。

ここで、前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成される。

また、前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とする。

さらに、前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより1つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行っている。

また、前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行っている。

また、前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うこととしている。

また、前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えている。

さらに、クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせている。

また、本発明の情報分類装置は、検索サービスで検索された複数の検索結果が入力され、入力された複数の検索結果に対してクラスタリング処理を施し、そのクラスタリング処理結果を出力するクラスタリングモジュールを含むものである。

また、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールをさらに備えることもできる。

また、前記クラスタリングモジュールによって得られたクラスタリング結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力するクラスタ順位設定モジュールとを有することもできる。

また、前記クラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールとを有することもできる。

また、本発明の記録媒体は、検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を出力する情報分類プログラムを記録した記録媒体であって、

その情報分類プログラムは、前記検索サービスからの検索結果を取得する手順と、

取得した検索結果に対してクラスタリング処理を施し、そのクラスタリング処

理結果を出力する手順とを含んでいるものである。

ここで、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行った後に、前記クラスタリング処理を行うこともできる。

また、前記クラスタリング処理によって得られたクラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力する手順とを含んでなるものである。

また、前記クラスタリング処理によって得られたクラスタリング処理結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成手順と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順とを含むものである。

発明を実施するための最良の形態

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明の情報分類方法および情報分類装置についての説明であるとともに、本発明の情報分類処理プログラムを記録した記録媒体における情報分類処理プログラムの具体的な処理内容をも含むものである。

(第1の実施形態)

図1は本発明の第1の実施形態を示すもので、大きく分けると、検索サービス1、変換モジュール2、クラスタリングモジュール3とから構成され、変換モジュール2とクラスタリングモジュール3が情報分類装置に相当する。

検索サービス1はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえばwebページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス1で検索された検索結果はファイル形式で出力され、クラスタリングモジュール3に渡されるが、検索サービス1は複数存在し、それ

その検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール2が設けられる。

クラスタリングモジュール3は、検索サービス1により出力された検索結果ファイル内容（変換モジュール1による変換後のファイル内容）からクラスタリングするために必要な情報（クラスタリング対象情報という）を抽出するクラスタリング対象情報抽出部31、そのクラスタリング対象情報として抽出された情報を形態素解析する形態素解析部32、その形態素解析結果に基づいてクラスタリングを行うクラスタリング処理部33などを有している。

クラスタリング対象情報抽出部31は、変換モジュール部2で変換された検索エンジン1の検索結果から、クラスタリング対象情報を抽出するものであり、このクラスタリング対象情報としては幾つか考えられる（後述する）。この実施の形態では、検索結果として抽出された多数の文書のそれぞれのタイトル（見出し）をクラスタリング対象情報として抽出するものとする。たとえば、検索結果として図2に示すような複数の文書D1、D2、・・・、D7が得られたとする。これら文書D1、D2、・・・、D7はタイトルT1、T2、・・・、T7と、それに対する本文A1、A2、・・・、A7を持っているものとする。

クラスタリング対象情報抽出部31は、このような検索結果に対し、それぞれの文書D1、D2、・・・、D7を解析し、それぞれの文書のタイトルを検出する。このクラスタリング対象情報抽出部31が行うタイトルの検出は、具体的には次のようにして行う。

まず、第1の方法として、文書構造様式によってタイトルと規定される部分があればその部分をタイトルとする。また、第2の方法として、文書構造様式によって、標準より大きな文字で表示する指定がなされている部分があれば、その部分をタイトルとする。また、第3の方法として、定められた数の文または単語を文書先頭より抽出し、その抽出した部分をタイトルとする。さらには、これら第1、第2、第3の方法を順次行い、第1の方法を行ったとき、タイトルと規定されている部分があればその部分をタイトルとし、タイトルと規定される部分が存在しなければ、第2の方法を行い、標準より大きな文字で表示する指定がなされ

ている部分があれば、その部分をタイトルとし、標準より大きな文字で表示する指定がなされていなければ、第3の方法を行ってタイトルを抽出する。

形態素解析部332は、クラスタリング対象情報抽出部331でそれぞれの文書からタイトルとして抽出された部分を形態素解析する。

クラスタリング処理部333は、図3に示すように、特徴要素抽出部331、特徴テーブル作成部332、文書分類部333、分類結果記憶部334、出力制御部335、表示部336などを有している。特徴要素抽出部331は、形態素解析部332で形態素解析された結果から特徴要素を抽出する。

特徴テーブル作成部332は、特徴要素抽出部331で抽出された特徴要素とそれぞれの文書D1～D7との関係を示す特徴テーブルを作成する。なお、この特徴テーブルの具体的な内容については後述する。

文書分類部333は、上述の特徴テーブルの内容を参照し、文書D1, D2, ..., D7を意味的に共通性のある複数のクラスタに分類する。つまり、文書D1, D2, ..., D7のそれぞれのタイトルT1, T2, ..., T7に存在する特徴要素に基づいて、共通する特徴要素を持つ文書を1つのまとまりとし、そのまとまりを1つのクラスタとする。なお、この文書分類部333は同義特徴辞書(図示せず)を有し、共通する特徴要素を持つ文書を1つのまとまりとする処理を行う際、共通する特徴要素であるか否かの判断を、その同義語辞書を用いて同義語が有るか否かにより行い、同義語が存在する場合にはそれを同じクラスタとするというようなことを行うことも可能である。

分類結果記憶部334は、文書分類部333によって分類された内容を記憶する。出力制御部335は分類結果記憶部334の内容を読み出して分類結果として表示部336に表示させる。

このような構成において、本発明の情報分類処理について説明する。本発明が行う情報分類処理手順は概略的には、図4のフローチャートに示すように、まず、汎用の検索エンジンで検索された検索結果を取得し(ステップS1)、取得した検索結果に対しクラスタリング処理を施し(ステップS2)、そのクラスタリング処理結果を出力する(ステップS3)。以下、具体例を参照しながら詳細に説明する。

ここでは、ユーザの入力したキーワードによって図2で示した文書D1, D2, ..., D7が検索結果として検索サービス1から出力されたとする。この検索結果はファイル形式で出力され、変換モジュール2でクラスタリングモジュール3が処理できる形式に変換されたのち、クラスタリングモジュール3に与えられる。

クラスタリングモジュール3に入力されたこれらの文書D1, D2, ..., D7は、クラスタリング対象情報抽出部3.1にてタイトルが抽出される。たとえば、文書D1についてはタイトルT1が検出され、文書D2についてはタイトルT2が検出され、文書D3についてはタイトルT3が検出されるというように、それぞれの文書D1, D2, ..., D7のタイトルT1, T2, ..., T7が抽出される。

そして、形態素解析部3.2によってそれぞれのタイトルT1, T2, ..., T7が形態素解析されたのち、その形態素解析結果がクラスタリング処理部3.3に与えられる。クラスタリング処理部3.3では、特徴要素抽出部3.3.1によって、形態素解析部3.2からの形態素解析結果に基づいて、それぞれのタイトルT1, T2, ..., T7に存在する特徴要素が抽出される。

そして、特徴テーブル作成部3.3.2により、それぞれの特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルが作成される。この特徴テーブルの例を図5に示す。なお、ここでは、文書数が3つ以上取り出される特徴要素とその特徴要素をタイトルに含む文書との関係を示し、特徴テーブル内に示される数値は、その特徴要素が各文書のタイトルに幾つ含まれているかの数を示している。たとえば、「用紙」という特徴要素は、文書D1, D4, D6, D7のタイトルT1, T4, T6, T7に、それぞれ1個ずつ含まれていることを示している。

図5の特徴テーブルからもわかるように、「用紙」という特徴要素をタイトルに含む文書は、文書D1, D4, D6, D7であり、また、「カセット」という特徴要素をタイトルに含む文書は、文書D1, D4, D7であり、さらに、「増設」という特徴要素をタイトルに含む文書は、文書D2, D3, D5, D7である。なお、先に説明した図2において、これら各特徴要素部分にはアンダーライ

ンが施されている。

そして、文書分類部 333 はこのような特徴テーブルを参照して、それぞれの特徴要素ごとのクラスタリングを行う。その分類結果を図 6 に示す。

このような分類結果は分類結果記憶部 334 に格納される。図 6 に示される分類結果において、たとえば、「用紙」で分類されたクラスタ（文書 D 1, D 4, D 6, D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 1 は用紙カセットについての内容であり、文書 D 4 は用紙設定についての内容であり、文書 D 6 は印刷された後の用紙の汚れについての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

このように、これらの文書 D 1, D 4, D 6, D 7 はどれも用紙に関する内容であり、1つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

また、「カセット」で分類されたクラスタ（文書 D 1, D 4, D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 1 は用紙カセットについての内容であり、文書 D 4 は用紙設定についての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

また、このように、これらの文書 D 1, D 4, D 6, D 7 にはどれも用紙をセットすることに関する内容が含まれており、1つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

また、「増設」で分類されたクラスタ（文書 D 2, D 3, D 5, D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 2 はメモリの増設についての内容であり、文書 D 3 はインタフェースカードの増設についての内容であり、文書 D 5 はハードディスクの増設についての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

このように、これらの文書 D 2, D 3, D 5, D 7 はどれも何かを増設する場合についての内容であり、1つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

このような適切な分類が行える理由としては、それぞれの文書のタイトルから特徴要素を抽出し、その特徴要素に基づいて文書を分類しているからである。つ

まり、文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることが多い。したがって、文書のタイトルに含まれる特徴要素を用いて分類を行うことにより、分類結果が散漫になることが少なく、また、ノイズクラスタが生成される率も少なくすることができる。また、各文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることから、文書の制作者側の視点による分類が得られる。

なお、ユーザに実際に提示されるクラスタリング結果は、たとえば図7に示すような内容であり、それぞれの特徴要素とその特徴要素を有するタイトルの一覧が提示される。そして、ユーザはこのようなクラスタリング結果の一覧を見て、自分の欲しい情報の入っていそうなタイトル部分をクリックすれば、そのタイトルに対応する本文が表示されるというような表示処理がなされる。

このように、この実施の形態では、ユーザがある汎用の検索サービスを利用し、その検索サービスに何らかのキーワードを入力することによって、複数の文書D1, D2, ..., D7が検索された場合、これら複数の文書D1, D2, ..., D7のタイトルT1, T2, ..., T7を抽出し、そのタイトルに含まれる特徴要素に基づいて、それぞれの文書D1, D2, ..., D7に対しクラスタリング処理を行う。

これによって、従来では、検索サービスによって検索された検索結果は単に羅列的に提示されるだけであったものが、本実施形態においては、検索された検索結果のタイトル内容に基づいたクラスタリング結果をユーザに提示できる。そのクラスタリング結果（図7参照）は、それぞれのタイトルに含まれる特徴要素ごとにそれぞれのタイトルが分類された内容となり、ユーザが見易いように整理されたものとなる。

そして、その中から見たい情報があれば、タイトル部分をクリックすれば、それぞれのタイトルを有する文書を表示させることができる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。

検索サービスはそれぞれに得意の分野があることも多く、たとえば、ある検索

サービスはスポーツ関係の情報を多数保有し、ある検索サービスは学術関係の情報を多数保有し、また、ある検索サービスは芸能関係の情報を多数保有しているというように、それぞれの得意の分野が存在する場合も多い。これらそれぞれの得意分野については豊富な情報を所有しており、ユーザの所望とする情報が適切に取り出される可能性が高い。したがって、情報検索を行う際は、検索サービスを使い分けることも普通に行われる。このように、複数の検索サービスを用いる場合のクラスタリング処理について以下に説明する。

図8は複数の検索サービスを用いてこれまで説明したクラスタリング処理を行うための情報分類装置の構成を説明する図であり、ここでは、複数の検索サービスとして第1の検索サービス1a、第2の検索サービス1b、第3の検索サービス1cの3つの検索サービスが存在するものとする。

このように複数の検索サービス（ここでは検索サービス1a、1b、1c）を対象とする場合には、それぞれの検索サービスにより検索された検索結果の内容、長さ、検索結果出力順序などがまちまちなので、それぞれの検索サービス1a、1b、1cからのファイルをクラスタリングモジュール3で処理可能な形式に変換する変換モジュール2a、2b、2cを検索サービス1a、1b、1cに対応して用意する。なお、クラスタモジュール3の構成は図1と同じであるので同一部分には同一符号が付されている。

このような構成であれば、ユーザは検索しようとする情報の分野に応じて検索サービスを使い分けることができる。たとえば、第1の検索サービス1aがスポーツ関係の情報検索に向いている検索サービスであるとすれば、スポーツ関係の情報を検索しようとする場合、第1の検索サービス1aを用いて検索を行う。また、第2の検索サービス1bが学術関係の情報検索に向いている検索サービスであるとすれば、学術関係の情報を検索しようとする場合、第2の検索サービス1bを用いて検索を行う。

このように、ユーザは検索しようとする情報によって検索サービスを選択することができ、しかも、それによって検索された結果は、クラスタリングモジュール3によってクラスタリング処理がなされ、ユーザが見やすいように整理された状態で提示できる。なお、このクラスタリング処理については、すでに説明した

のでここではその説明は省略する。

このように複数の検索サービスを選択的に使用できるようにすれば、それぞれの得意分野に応じた検索が可能となるばかりでなく、ある1つの検索サービスが混み合っているような場合には、他の検索サービスに切り換えて検索を行うというような柔軟な検索も可能となる。

さらに、複数の検索サービスによる検索処理を並行に行い、それぞれの検索サービスによる検索結果を1つにまとめて、まとめられた内容についてクラスタリング処理することも可能である。これを図9により簡単に説明する。

図9の構成は図8の構成において、第1～第3の検索サービス1a, 1b, 1cに対応して設けられた変換モジュール2a, 2b, 2cとクラスタリングモジュール3との間に、それぞれの検索結果を収集して1つにまとめる検索結果収集部4を設けた構成となっている。その他は、図8と同じ構成であり、同一部分には同一符号が付されている。

このような構成であれば、複数の検索サービス（ここでは第1～第3の検索サービス1a, 1b, 1c）では、ユーザの入力したキーワードに対して並行して検索処理を行い、それぞれの検索サービス1a, 1b, 1cによって検索されたそれぞれの検索結果を、それぞれの検索サービス1a, 1b, 1cに対応する変換モジュール部2a, 2b, 2cでクラスタリングモジュール3が処理可能な形式に変換し、それぞれの変換後のファイルを検索結果収集部4に与え、それぞれの検索結果を1つにまとめる。そして、そのまとめられた検索結果は、クラスタリングモジュール3に入力され、これまで説明したようなクラスタリング処理を行う。

このように、複数の検索サービスを用いて検索処理を行うことにより、1つの検索サービスでは検索できなかった幅広い情報を取得することも可能である。また、検索範囲が広がるので、網羅的に情報を探すことができ、どのような情報が世の中に存在するのかを見極めるのに非常に有効となる。このようにして得られた情報は、これまで説明したようなクラスタリング処理がなされ、見やすい形でユーザに提示される。

また、この図9に示した第1～第3の検索サービス1a, 1b, 1cを用いて

検索処理を並行して行うような場合、それぞれの検索サービス 1 a, 1 b, 1 c で得られた検索結果（変換モジュール 2 a, 2 b, 2 c 出力）を 1 つにまとめたあと、クラスタリング処理を行うのではなく、それぞれの検索サービス 1 a, 1 b, 1 c で得られた検索結果（変換モジュール 2 a, 2 b, 2 c 出力）それぞれについてクラスタリング処理を行い、それぞれのクラスタ処理結果をユーザに提示するようにしてもよい。

これによって、広範囲にわたって存在する情報を効率よく検索することが可能となり、また、ユーザは第 1～第 3 の検索サービス 1 a, 1 b, 1 c によるそれぞれの検索結果のそれぞれのクラスタリング結果を見比べたりすることができ、それぞれの検索サービスの特色などを知ることできる。

なお、本実施形態は、上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、上述の実施の形態では、クラスタリングを行うための情報（クラスタリング対象情報）として、検索されたそれぞれの文書のタイトルを用いた例について説明したが、これは、タイトルだけでなく、たとえば、URL アドレス（http://を取り除いた部分）、更新日時（単純な時間または最近 1 カ月以内の更新日時）、ファイルサイズ（web ページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、これらのどれを用いるかは、最初にメニューなどで選択項目を選ぶことで可能となる。また、選んだ項目が無い場合には、他の項目を代用する。たとえば、タイトルを選んだ場合、web ページにタイトルが無い場合には、URL アドレスを代用する。

また、本実施形態における情報分類処理を行う処理プログラムは、フロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録させておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてもよい。

（第 2 の実施形態）

次に、本発明の第2の実施形態について説明する。

第1の実施形態の説明において述べたように、文書のタイトルから特徴要素を抽出してクラスタリングする手法は、演算量や処理時間の面で優れ、適切なクラスタリングが可能となるが、クラスタリングを行うための情報量は文書全体から見れば少ないので、全てが適切にクラスタリングされるとは限らない。特に、タイトルが文書の内容を適切に表していなかったり、文書内容とは大きくかけ離れた奇抜なタイトルが付けられていたりする場合が考えられる。このような場合には、クラスタリング精度は大きく低下し、良好なクラスタリング結果は得られないことになる。

また、特徴要素を抽出して、その特徴要素に基づいてクラスタリングする手法は、特徴要素の頻度などを調べ、それによって、機械的に文書を分類してクラスタリングするものである。このようなクラスタリングでは、文書の意味を解析しているわけではないので、得られたクラス（クラスタリングされることによって得られる文書群の1つの集合）が必ずしも意味的な共通性のある文書の集合となるとは限らない。

情報分類においては、以上のような場合であっても、ユーザの検索要求に沿ったクラスタリング結果を提示できるようにすることが好ましい。

本実施形態においては、汎用の検索サービスで得られた検索結果をクラスタリング処理を施し、かつ、クラスタリングによって得られたクラスの順位を再構成して表示することで、ユーザの検索要求に沿ったクラスタリング結果を提示できるようにしている。

以下に、この第2の実施形態について詳細に説明する。

図10は、第2の実施形態をの装置構成を示すもので、大きく分けると、検索サービス101、変換モジュール102、クラスタリングモジュール103、クラス順位再構成モジュール104とから構成され、変換モジュール102、クラスタリングモジュール103、クラス順位再構成モジュール104が情報分類装置に相当し、特に、クラス順位再構成モジュール104を設けた点に本実施形態の特徴がある。

検索サービス101はインターネットなどで一般に広く使用されている汎用の

検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえば web ページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス 101 で検索された検索結果はファイル形式で出力され、クラスタリングモジュール 103 に渡されるが、複数の検索サービスが存在する場合、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール 102 が設けられる。

クラスタリングモジュール 103 は、検索サービス 101 により出力された検索結果（変換モジュール 102 による変換後のファイル内容）に対しクラスタリング処理を行うもので、この実施の形態では、それぞれの文書から文書のタイトルを抽出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいてクラスタリング処理を行う。

具体的には、それぞれの文書からタイトルとして抽出された部分を形態素解析し、形態素解析された結果から特徴的な用語を特徴要素として抽出する。その後、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成する。この特徴テーブルは、たとえば、抽出されたそれぞれの特徴要素が各文書のタイトルに幾つ含まれているかを、それぞれの特徴要素とそれぞれの文書と対応づけて示すもので、一例として、それぞれの文書のタイトルから、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」というような特徴要素が抽出されたとすると、これらの特徴要素が、それぞれの文書のそれぞれのタイトルに、それぞれ何個含まれているかを示す内容となっている。

このような特徴テーブルに基づいて、多数の文書を意味的に共通性のある複数のクラスタに分類する。つまり、それぞれの文書のそれぞれのタイトルに存在する特徴要素に基づいて、共通する特徴要素を持つ文書を 1 つのまとまりとし、そのまとまりを 1 つのクラスタとする。

このクラスタリングモジュール 103 から、たとえば、図 11 のようなクラスタリング結果が出力されたとする。この図 11 は、前述したように、クラスタリングされることによって得られた各クラスタの名称（上述の特徴要素に相当し、それをここではクラスタ名と呼んでいる）と、そのクラスタに属するそれぞれの

文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、前述したように、それぞれのタイトルごとにスコアを示す数値などが一覧表形式で示されている。

なお、このスコアは、前述したように、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、ここでは、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるとする。このスコアはキーワードとの適合度を表すものであるので、その単位としては、%や点数など検索サービスによって異なるがこの実施の形態では点数で表すものとする。

そして、このクラスタリングモジュール103によってクラスタリングされた段階のクラスタリング結果は、図11に示されるように、クラスタの配置はそれぞれのクラスタに含まれる文書の数の多い順となっている。前述したように、この場合、上から順に、概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタの順となっている。

クラスタ順位再構成モジュール104は、クラスタリングモジュール103で出力されたクラスタリング結果に基づいて、それぞれのクラスタの表示順位を再構成するもので、その処理内容の詳細については後に説明する。

このような構成において、本発明の第2の実施形態について説明する。

本実施形態において行う情報分類処理手順は概略的には、図12のフローチャートに示すように、まず、汎用の検索サービス101で検索された検索結果を取得し（ステップ12S1）、取得した検索結果に対しクラスタリング処理を施し（ステップ12S2）、そのクラスタリング結果を出力する（ステップ12S3）。そして、そのクラスタリング結果に対し、それぞれのクラスタ順位の再構成を行い（ステップ12S4）、再構成されたクラスタリング結果を出力する（ステップ12S5）。以下、具体例を参照しながら詳細に説明する。

この実施の形態では、クラスタリングモジュール103が行うクラスタリング処理は、検索サービス101で検索された文書に対し、それぞれの文書のタイトルを抽出し、そのタイトルから特徴要素を抽出して、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成して、その特徴テーブルの内容に基づいて、それぞれの文書を意味的に共通性のある複数のクラスタに分類する。

また、この実施の形態では、ユーザが「半導体」というキーワードを検索要求として検索サービス101に入力し、それによって得られた多数の文書がクラスタリングモジュール103によってクラスタリングされ、そのクラスタリング結果が図11に示すような結果であったとする。

このクラスタリングモジュール103からのクラスタリング結果は、クラスタ順位再構成モジュール104に入力され、以下に示すような処理がなされる。

まず、図11で示されるクラスタリング結果における各クラスタ（概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタ）において、それぞれのクラスタに含まれる文書対応に付されたスコアを利用して、そのスコアの値の平均を求める。この場合、それぞれのクラスタごとにスコアの値を足し算し、その足し算して得られた結果をそのクラスタに含まれる文書数で割る単純平均を求める。

たとえば、概要クラスタで考えると、この図11に示す検索結果においては、そのクラスタ内のスコアの合計が579点あって、文書数が16個であるので、平均のスコアは約36点と求められる。また、「LP」クラスタで考えると、そのクラスタ内のスコアの合計が450点であって、文書数が16個であるので、平均のスコアは約28点と求められる。同様に、「仕様」クラスタは、そのクラスタ内のスコアの合計が413点であって、文書数が14個であるので、平均のスコアは約29点と求められ、「デバイス」クラスタは、そのクラスタ内のスコアの合計が849点であって、文書数が9個であるので、平均のスコアは約94点と求められ、「半導体」クラスタは、そのクラスタ内のスコアの合計が757点であって、文書数が7個であるので、平均のスコアは約108点と求められ、「電子」クラスタは、そのクラスタ内のスコアの合計が349点であって、文書数が4個であるので、平均のスコアは約87点と求められる。

以上のようにして計算された平均のスコアを各クラスタのスコア（クラスタスコアと呼ぶ）とする。そして、このクラスタスコアの高い順にクラスタの順位を再構成する。

すなわち、この場合、クラスタスコアの最も高いクラスタは、半導体クラスタの108点であり、第2位はデバイスクラスタの94点であり、第3位は電子ク

ラストの87点であり、以下、概要クラスタ(36点)、仕様クラスタ(29点)、LPクラスタ(28点)といった順序となる。

このようにして、それぞれのクラスタごとにクラスタスコアを計算し、求められたクラスタスコアの高い順にクラスタ順位を再構成する。

この再構成されたクラスタリング結果を一覧表形式で表したものが図13である。図13によれば、表の最上段に半導体クラスタが位置し、2番目にデバイスクラスタ、3番目に電子クラスタ、以下、概要クラスタ、仕様クラスタ、LPクラスタといった順序となる。この図13のクラスタリング結果によれば、ユーザの入力した「半導体」というキーワードに対し、そのキーワードに適合する文書が多く含まれるクラスタが上位に来ていることがわかる。

この図13のクラスタリング結果と図11のクラスタリング結果を比較すると、図11のクラスタリング結果では、ユーザの入力した「半導体」というキーワードに対し、そのキーワードとは直接には関係しないような文書で構成される概要クラスタ、LPクラスタ、仕様クラスタといったクラスタが上位に位置し、キーワードに大きく関係するような文書が含まれると思われる半導体クラスタ、デバイスクラスタ、電子クラスタといったクラスタが下位に位置しているが、図13では、それが逆転し、キーワードに大きく関係するような文書が含まれると思われるクラスタが上位に位置するようになる。

なお、クラスタスコアが同じ値となった場合には、クラスタ内に含まれる文書数の多い方を上位とするなどの措置を講ずる。

ここで、各クラスタ内のスコアの合計及び平均については、図13のように表示しても良いし、また、表示しなくても良い。

以上説明したように、単純にそれぞれのクラスタに含まれる文書数(1つのクラスタにまとめられた文書数)によって順位付けするのではなく、それぞれのクラスタごとにそのクラスタに含まれる文書に付されたスコアに基づいてクラスタの順位を決めることによって、キーワードに適合したクラスタ順位が得られる。

なお、図13に示すようなクラスタリング結果がユーザに表示され、ユーザはこのようなクラスタリング結果の一覧表を見て、自分の欲しい情報の入っている文書のタイトル部分をクリックすれば、そのタイトルに対応する本文が表示

されるというような表示処理がなされる。

以上説明したように、本発明の第2の実施形態では、ユーザの入力したキーワードによって検索された多数の文書に対し、これら多数の文書のタイトルに含まれる特徴要素に基づいてクラスタリング処理し、さらに、そのクラスタリング結果に対して、それぞれのクラスタごとにそのクラスタに含まれる文書のスコアの平均を求める。そして、その平均のスコアをクラスタスコアとし、それぞれのクラスタごとのクラスタスコアに基づいて、クラスタ順位の再構成を行う。つまり、クラスタスコアの大きい順にクラスタの並べ替えを行い、図13に示すようなクラスタリング結果として表示する。

これによって、ユーザの欲しい情報の入っていそうなクラスタが一覧表の上位に位置した状態で表示されているので、自分の欲しい情報を探しやすくなる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。

検索サービスはそれぞれに得意の分野があることも多く、たとえば、ある検索サービスはスポーツ関係の情報を多数保有し、ある検索サービスは学術関係の情報を多数保有し、また、ある検索サービスは芸能関係の情報を多数保有しているというように、それぞれの得意の分野が存在する場合も多い。これらそれぞれの得意分野については豊富な情報を所有しており、ユーザの所望とする情報が適切に取り出される可能性が高い。したがって、情報検索を行う際は、検索サービスを使い分けることも普通に行われる。

このように、複数の検索サービスを用いてクラスタリング処理する場合には、それぞれの検索サービスにより検索された検索結果の内容、長さ、検索結果出力順序などがまちまちなので、それぞれの検索サービスからのファイルをクラスタリングモジュール103で処理可能な形式に変換する変換モジュール102を複数の検索サービスに対応して用意する。そして、さらに、そのクラスタリング結果におけるクラスタ順位再構成を行う場合には、それぞれのクラスタのクラスタスコアを求める処理をそれぞれの検索サービスに対応して行うようにする。

たとえば、本実施形態のクラスタ順位再構成処理についていえば、検索サービ

スによって幾つかの対策を講じる必要がある。たとえば、スコアの分布の幅が非常に大きい場合（たとえば、スコアを表す数値が最大1000から最小は2など）は、対数を取って計算するなどの措置を講じたり、また、きわめてスコアの値が小さい文書（たとえば、殆どの文書が数百のスコアの値があるのに2や3の値しかない文書）はクラスタリング対象から外すといった措置を講じる。

このように、複数の検索サービスに対応できるようにすることで、ユーザは検索しようとする情報の分野に応じて検索サービスを使い分けることができ、それぞれの得意分野に応じた検索が可能となるばかりでなく、ある1つの検索サービスが混み合っているような場合には、他の検索サービスに切り換えて検索を行うというような柔軟な検索も可能となる。

なお、本発明の第2の実施形態は、上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、これまで説明した実施形態では、それぞれのクラスタのクラスタスコアは、そのクラスタに含まれる文書のスコアの単純平均を用いた例について説明したが、このクラスタスコアとしては、それぞれのクラスタ内に含まれる文書のなかで最大のスコアを有する文書のスコアを用いるようにしてもよく、また、それぞれのクラスタ内に含まれる文書に付されたスコアのなかで中央に位置する文書のスコアを用いるようにしてもよい。

このように、クラスタごとのスコアの最大値を用いることで、クラスタ順位を決めるために足し算したり割り算したりという計算を行う必要がなく、計算量を少なくすることができ、しかも、同じクラスタ内に、極端に低いスコアを持つ文書が少数あっても、その影響をあまり受けないようにすることができる。また、クラスタごとのスコアの中央値を用いる場合も、最大値を用いるのと同様、クラスタ順位を決めるための計算量を少なくすることができ、しかも、中央値の場合、同じクラスタ内に、極端に高いスコアや極端に低いスコアを持つ文書が少数あっても、その影響をあまり受けないようにすることができる。

また、本実施形態では、クラスタリングを行うための情報（クラスタリング対象情報）として、検索されたそれぞれの文書のタイトルを用いた例について説明したが、これは、タイトルだけでなく、たとえば、URLアドレス（http://を取

り除いた部分）、更新日時（単純な時間または最近 1 カ月以内の更新日時）、ファイルサイズ（web ページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、これらのどれを用いるかは、最初にメニューなどで選択項目を選ぶことで可能となる。また、選んだ項目が無い場合には、他の項目を代用する。たとえば、タイトルを選んだ場合、web ページにタイトルが無い場合には、URL アドレスを代用する。

（第 3 の実施形態）

次に、本発明の第 3 の実施形態について説明する。

情報分類処理において、クラスタリングによって得られたクラスタの数がそれほど多くはない場合、ユーザはそのクラスタリング結果全体を把握するのにそれほど多くの時間を費やさずに済む。

しかし、クラスタリングによって得られたクラスタの数は時として何十個あるいは何百個といった膨大な数量となる場合もある。このような場合、クラスタリング結果全体を見るだけでも大変である。

そこで本発明の第 3 の実施形態においては、汎用の検索サービスで得られた検索結果をクラスタリング処理を施し、かつ、クラスタリングによって得られたクラスタリング結果全体の概要を一目で把握できるような一覧表を作成することで、ユーザが自分の欲しい情報を効率よく探すことができるようにしている。

以下に、第 3 の実施形態について詳細に説明する。

図 14 は本発明の第 3 の実施形態の概略の構成を示すもので、大きく分けると、検索サービス 141、変換モジュール 142、クラスタリングモジュール 143、クラスタリング結果概要一覧表作成モジュール（以下、概要一覧表作成モジュールという）144、表示制御モジュール 145 とから構成され、変換モジュール 142、クラスタリングモジュール 143、概要一覧表作成モジュール 144、表示制御モジュール 145 が情報分類装置に相当し、特に、概要一覧表作成モジ

ジュール 1 4 4 を設けたことに本発明の第 3 の実施形態の特徴がある。

検索サービス 1 4 1 はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえば web ページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス 1 4 1 で検索された検索結果はファイル形式で出力され、クラスタリングモジュール 1 4 3 に渡されるが、複数の検索サービスが存在する場合、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール 1 4 2 が設けられる。

クラスタリングモジュール 1 4 3 は、検索サービス 1 4 1 により出力された検索結果（変換モジュール 1 4 2 による変換後のファイル内容）に対しクラスタリング処理を行うもので、この実施の形態では、それぞれの文書から文書のタイトルを抽出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいてクラスタリング処理を行う。

具体的には、それぞれの文書からタイトルとして抽出された部分を形態素解析し、形態素解析された結果から特徴的な用語を特徴要素として抽出する。その後、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成する。この特徴テーブルは、たとえば、抽出されたそれぞれの特徴要素が各文書のタイトルに幾つ含まれているかを、それぞれの特徴要素とそれぞれの文書と対応づけて示すもので、一例として、それぞれの文書のタイトルから、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」というような特徴要素が抽出されたとすると、これらの特徴要素が、それぞれの文書のそれぞれのタイトルに、それぞれ何個含まれているかを示す内容となっている。

このような特徴テーブルに基づいて、多数の文書を意味的に共通性のある複数のクラスタに分類する。つまり、それぞれの文書のそれぞれのタイトルに存在する特徴要素に基づいて、タイトルの中に共通する特徴要素を持つ文書を 1 つのまとまりとし、そのまとまりを 1 つのクラスタとする。

このクラスタリングモジュール 1 4 3 から、たとえば、第 2 の実施形態において説明した図 1 1 のようなクラスタリング結果が出力されたとする。この図 1 1

は前述したように、クラスタリングされることによって得られた各クラスタの名称（上述の特徴要素に相当し、それをここではクラスタ名と呼んでいる）と、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

なお、このスコアは、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、ここでは、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるとする。

そして、このクラスタリングモジュール 143 によって得られたクラスタリング結果におけるそれぞれのクラスタの配列は、図 11 に示されるように、ここでは、それぞれのクラスタに含まれる文書の数の多い順となっている。前述したように、この場合、上から順に、概要クラスタ、LP クラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタの順となっている。

概要一覧表作成モジュール 144 は、クラスタリングモジュール 143 から出力されたクラスタリング結果に基づいて、それぞれのクラスタのクラスタ名に基づいて、クラスタリング結果の概要を表すクラスタリング結果概要一覧表（概要一覧表という）を作成するものである。

表示制御モジュール 145 は、クラスタリングモジュール 143 によって得られたクラスタリング結果や、概要一覧表作成モジュール 144 で作成された概要一覧表をユーザに提示するために様々な表示制御を行うもので、この実施の形態では、クラスタリング結果とともに概要一覧表を表示させることは勿論、クラスタリング結果と概要一覧表のリンク部分の相互の表示制御、さらには、注目すべきクラスタをユーザの目に付きやすい表示とするといった様々な表示制御を行う。その具体的な表示制御内容については後述する。

このような構成において、本発明の第 3 の実施形態の情報分類処理について説明する。本実施形態において行う情報分類処理手順は概略的には、図 15 のフローチャートに示すように、まず、汎用の検索サービス 1 で検索された検索結果を取得し（ステップ 15 S1）、取得した検索結果に対しクラスタリング処理を施し（ステップ 15 S2）、そのクラスタリング結果を出力する（ステップ 15 S

3)。そして、そのクラスタリング結果に基づいて概要一覧表作成処理を行い(ステップ15S4)、作成された概要一覧表を上述のクラスタリング結果とともに表示する(ステップ15S5)。なお、この作成された概要一覧表を上述のクラスタリング結果とともに表示する際の表示の仕方としては、たとえば、クラスタリング結果に概要一覧表を画面上で重ねて表示するようにしてもよく、また、概要一覧表とクラスタリング結果を別個に配置し、概要一覧表の後に続いてクラスタリング結果を表示させるようにしてもよい。なお、クラスタリング結果の容量が多い場合には、画面をスクロールすることによって、クラスタリング結果を順次画面上に現すようにする。

以下、本発明の第3の実施形態における情報分類処理内容について具体例を参照しながら詳細に説明する。

この実施の形態では、クラスタリングモジュール143が行うクラスタリング処理は、検索サービス141で検索された文書に対し、それぞれの文書のタイトルを抽出し、そのタイトルから特徴要素を抽出して、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成して、その特徴テーブルの内容に基づいて、それぞれの文書を意味的に共通性のある複数のクラスタに分類する。また、この実施の形態では、ユーザが「半導体」というキーワードを検索要求として検索サービス141に入力し、それによって得られた多数の文書がクラスタリングモジュール143によってクラスタリングされ、そのクラスタリング結果が図11に示すような結果であったとする。

このクラスタリングモジュール143からのクラスタリング結果は、概要一覧表作成モジュール144に入力され、以下に示すような処理がなされる。

まず、図11で示されるクラスタリング結果における各クラスタ(概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタ)において、それぞれのクラスタ名(「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」)を用いた概要一覧表を作成し、その概要一覧表をクラスタリング結果とともに表示する。

図16は概要一覧表1610をクラスタリング結果1620とともに表示した例を示すもので、この表示例では、概要一覧表1610に続いてクラスタリング

結果1620を表示した例である。ここで用いたクラスタリング結果1620は、クラスタ数が6個ときわめて少ない例であるが、実際には、クラスタ数は何十個あるいは何百個というような場合もあり、その中から自分の欲しい情報を探すには、クラスタリング結果全体を見ながら探すことになるので、単に、クラスタリング結果全体を見せられただけでは自分の欲しい情報を探すのに大変な労力を要する。これに対し、クラスタ名による概要一覧表が表示されることによって、その概要一覧表を構成するクラスタ名を見渡すことで、クラスタリング結果にはどのようなクラスタがあるのか、自分の欲しい情報はその中のどのクラスタに入っているのかといったことをおおよそ見当付けることができる。

そして、概要一覧表1610を構成するそれぞれのクラスタ名とクラスタリング結果とはそれぞれ対応する部分がリンクされている。したがって、たとえば、クラスタリング結果のクラスタ数が多く、一つの画面上ですべてのクラスタリング結果が表示しきれないような場合でも、図11で示される概要一覧表1610の任意のクラスタ名をクリックすると、それに対応するクラスタリング結果1620のクラスタ部分が即座に表示され、その状態で、今度はクラスタリング結果のクラスタ名をクリックすると概要一覧表に即座に戻るといった表示制御が可能となる。

このような表示制御を行う際、使い易さをより一層向上させるために次のような機能を付加する。

まず、概要一覧表におけるクラスタ名の配列順序は、クラスタリング結果におけるそれぞれのクラスタの配列順序に従う。すなわち、図11のクラスタリング結果を例に取れば、そのクラスタリング結果におけるそれぞれのクラスタの配列順序は、それぞれのクラスタに含まれる文書数の多い順となっている。たとえば、概要クラスタには文書数は16個、LPクラスタにも文書数は16個、仕様クラスタには文書数は14個、デバイスクラスタには文書数は9個、半導体クラスタには文書数は7個、電子クラスタには文書数は4個といった具合である。したがって、この場合、概要一覧表においても、図11に示すように、一覧表の左から右に、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」の順とした配列とする。

なお、クラスタリング結果におけるクラスタ順位の付け方としては、文書数だけではなく、それぞれのクラスタのスコアに基づいて決めることもできる。図11のクラスタリング結果を見ると、前述したように、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

このスコアは、与えられたキーワードに対しその検索サービスが独自の方法で検索したとき、検索された文書対応に付けられた値であり、一般には、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、汎用の検索サービスによる検索結果にはこのスコアが付されるのが普通である。

このスコアは、検索方式により計算の仕方や値の考え方などが異なるものの、一般に、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるといえる。

したがって、それぞれのクラスタごとにそれぞれのクラスタにおけるスコアの平均などを求め、その平均のスコアの大きいクラスタ順に配列することも考えられる。このように、クラスタリング結果におけるそれぞれのクラスタ順位を、平均のスコアの大きいクラスタ順とした場合には、それによって作成される概要一覧表のクラスタ名の順序もそれに従う。

たとえば、概要クラスタの平均のスコア（このスコアはキーワードとの適合度を表すものであるので、その単位としては、%や点数など検索サービスによって異なるがこの実施の形態では点数で表すものとする）が約36点と求められ、LPクラスタの平均のスコアが約28点と求められ、仕様クラスタの平均のスコアが29点と求められ、デバイスクラスタの平均のスコアが約94点と求められ、半導体クラスタの平均のスコアが108点と求められ、電子クラスタの平均のスコアが87点と求められたとする。このように計算された平均のスコアを各クラスタのスコア（クラスタスコアと呼ぶ）とする。そして、このクラスタスコアの高い順にクラスタの順位を再構成すると、この場合、クラスタスコアの最も高いクラスタは、半導体クラスタの108点であり、第2位はデバイスクラスタの94点であり、第3位は電子クラスタの87点であり、以下、概要クラスタ（36点）、

仕様クラスタ（29点）、LPクラスタ（28点）といった順序となる。

このように、クラスタリング結果におけるそれぞれのクラスタの配列順序が、それぞれのクラスタのクラスタスコアの高い順となっているとすれば、それによって作成される概要一覧表のクラスタ名の順序も、クラスタリング結果のクラスタ配列順序に従って、「半導体」、「デバイス」、「電子」、「概要」、「仕様」、「LP」の順序とする。

ここで、各クラスタ内のスコアの合計及び平均については、図16のように表示しなくても良いし、また、表示しても良い。

一般に、クラスタリング結果のクラスタ順序は何らかの意味を持っているので、概要一覧表のクラスタ名の順序もそれに従うようにした方が、ユーザが情報を探す上で都合がよい場合が多い。ユーザは一般に上から順に見て行く傾向にあるので、概要一覧表のクラスタ名の配置順序をクラスタリング結果におけるクラスタの配置順序と同じにすれば、自分の欲しい情報を探すのに便利である。

次に、概要一覧表1610の或るクラスタ名をクリックしてそれにリンクされたクラスタリング結果1620のクラスタ部分を表示させる際、そのクラスタを囲っている枠（クラスタ枠という）の最上部を先頭にした表示を行う。これは、もしクラスタのクラスタ名部分を基準にしてそのクラスタ名部分を先頭にした表示を行うと、表示させたときに、そのクラスタ名に対応するクラスタ要素（図11ではクラスタに含まれる文書のタイトル）が表示画面上に表示されない場合があるという不都合が生じる。たとえば、画面上にクラスタリング結果1620の半導体クラスタが現れていない状態で、概要一覧表1610の「半導体」というクラスタ名をクリックされたとすると、そのクリック動作とほぼ同時にクラスタリング結果1620の半導体クラスタ部分の表示に移行することになるが、そのとき、そのクラスタ枠内の最上段の行に位置するクラスタ要素（この場合は、「157 半導体事業部環境方針」という文書のタイトル）が表示画面上で見えないということがしばしば生じる。

これを防ぐために、クラスタ枠を基準にしてそのクラスタ枠の最上部を先頭にした表示を行うようにする。これによって、そのクラスタ枠内の最上段に位置する行のクラスタ要素が確実に表示がなされるようになる。

また、さらに確実性を高めるため、それより1つ前に位置するクラスタの最終行のクラスタ要素を先頭にした表示を行うようにしてもよい。たとえば、前述の例で考えれば、概要一覧表1610の「半導体」というクラスタ名をクリックされたとすると、クラスタリング結果の半導体クラスタが表示されることになるが、このとき、それよりも1つ前のデバイスクラスタの最終行のクラスタ要素（図11では、「56 デバイス-半導体-ASSP」）といったクラスタ要素を先頭にした表示を行うようにしてもよい。このように、1つ前に位置するクラスタの最終行のクラスタ要素から表示させるようにすれば、本来表示すべきクラスタのクラスタ要素は確実に表示がなされるようになる。

次に、概要一覧表1610に表示されるそれぞれのクラスタ名は、クラスタリング結果1620におけるそれぞれのクラスタの内容によって表示の大きさや色を変える。このクラスタリング結果におけるそれぞれのクラスタの内容というのは、具体的には、それぞれのクラスタの重要度を示すもので、ユーザの与えたキーワードに対する適合度などであり、それぞれのクラスタに含まれる文書数やそれぞれのクラスタのスコアなどを用いて決める。たとえば、前述したように、それぞれのクラスタにおけるスコアの平均を求め、その平均のスコアの大きいクラスタが重要度の高いクラスタであるとすれば、重要度の高いクラスタに対する概要一覧表1610のクラスタ名の表示の仕方を変える。

たとえば、前述した例によれば、図11のクラスタリング結果におけるそれぞれのクラスタのなかで、半導体クラスタのクラスタスコアが最も高い値であったので、半導体クラスタに対応する概要一覧表のクラスタ名「半導体」の表示の仕方を他のクラスタのクラスタ名と異ならせる。具体的には、「半導体」というクラスタ名を他のクラスタ名と色を変えて表示したり、「半導体」というクラスタ名を囲っている枠だけを他よりも太くしたり、その枠内の面積を他よりも大きな面積としたり、さらにはそのクラスタ名をブリンク表示させたりすることによって、ユーザの目につきやすい表示を行う。

また、それぞれのクラスタにおいてそれぞれのクラスタに含まれる文書数の多いクラスタについても同様に、ユーザの目につきやすい表示を行うようにすることもできる。さらに、スコアの高いクラスタは色を他のクラスタとは異ならせ、

文書数の多いクラスタは枠内の面積を大きくするなど、クラスタの特徴によって表示の仕方を多種多様に設定するようにしてもよい。このようにすれば、どのクラスタがどの同様な特徴を持ったクラスタであるかということが、概要一覧表 1610 を見るだけで一目でわかる。

また、概要一覧表 1610 においてユーザの入力したキーワードが含まれるクラスタ名は、それをユーザに知らせるために他のクラスタ名と異なった表示を行う。

たとえば、図 11 のクラスタリング結果の例では、ユーザの与えたキーワードは「半導体」であるため、クラスタリング結果により得られたクラスタのうち、半導体クラスタはまさにキーワードそのものを含むクラスタである。

この場合、そのクラスタリング結果 1620 により作成された概要一覧表 1610 において、「半導体」部分は他のクラスタとは異ならせた表示として、たとえば、ブリンクさせたり、色を変えたり、両者を組み合わせるなど、ユーザの目に付きやすい表示を行う。一般に、ユーザは自分の入力したキーワードと同じものがあるかを探すことが多い。したがって、概要一覧表 1610 内に自分の与えたキーワードと同じクラスタ名が有るかどうかを一目でわかるようにすれば、ユーザにとっては自分の欲しい情報を探す上で便利なものとなる。

また、画面上にクラスタリング結果 1620 の或る部分が表示されている状態のときに、その表示内容から或るクラスタ名をクリックすると概要一覧表 1610 の表示に戻るが、そのとき、概要一覧表 1610 の先頭に戻って、概要一覧表 1610 の先頭部分からの表示がなされるようにする方が都合がよい。

概要一覧表 1610 はクラスタリング結果の概要を簡略化して表すものではあるが、場合によってはそのサイズがきわめて大きなものとなることもある。また、複数の概要一覧表が作成される場合もある。すなわち、これまでの説明では、それぞれの文書のタイトルを用い、そのタイトルによってクラスタリング処理しそれによって得られたクラスタリング結果に基づいて概要一覧表を作成する例について説明したが、クラスタリング処理は、タイトルだけでなく、URL アドレス (<http://>を取り除いた部分) などの情報を用いても行うことができる。

たとえば、図 11 のクラスタリング結果を得るために用いた多数の文書に対し、

URLを用いてクラスタリング処理し、それによって得られたクラスタリング結果に基づいて概要一覧表を作成することもできる。

図17は図11と同じ文書に対し、URLアドレスによってクラスタリングされたクラスタリング結果1730とそれに基づいて作成された概要一覧表1740の例を示すものである。この場合、クラスタリングされることによって得られたクラスタのクラスタ名としては、クラスタリングする際の用いられたそれぞれのURLアドレス「www.epson.co.jp」と「www.i-love-epson.ne.jp」や「その他」であり、概要一覧表1740を構成するクラスタ名は、これらのURLアドレス「www.epson.co.jp」と「www.i-love-epson.ne.jp」さらに「その他」が用いられる。

この例で説明したように、クラスタリング処理は、幾つもの方法によってなされる可能性があり、その場合、それぞれによって得られたクラスタリング結果に対する複数の概要一覧表が作成されることになる。

このように、複数の概要一覧表が作成される場合や、一つの概要一覧表であっても一覧表のデータ量がきわめて大きい場合、画面上でクラスタリング結果を見ていて、それに対応する概要一覧表のクラスタ名部分に戻したとき、対応するクラスタ名部分からの表示を行うのではなく、複数の概要一覧表が存在する場合には、複数の概要一覧表のなかで先頭に位置する概要一覧表の先頭部分からの表示がなされるようにする。また、1つの概要一覧表の場合でも、その概要一覧表の先頭部分からの表示がなされるようにする。

これは、たとえば、ユーザがクラスタリング結果を見ていて、もう一度、概要一覧表に戻ってクラスタリング結果全体を見たいというような場合に対処するものである。このような場合、概要一覧表の対応するクラスタ名部分に戻ると、戻った位置が一覧表全体のどの位置かがわからなくなる可能性があるが、概要一覧表が複数存在していても、あるいは概要一覧表のデータ量が多くても、概要一覧表全体の先頭部分からの表示がなされるようにすれば、一覧表全体を見渡すことができる。

以上説明したような種々の機能を追加することによって、概要一覧表の有用性をより一層高めることができる。

以上説明したように、この実施の形態では、膨大な情報をクラスタリングし、そのクラスタリング結果の概要が一目でわかるような概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体を見渡すことができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、その概要一覧表とクラスタリング結果との間に前述したような種々の機能を持たせることによって、ユーザの与えたキーワードに対し重要度の高いクラスはどれであるかなどが一目でわかり、また、概要一覧表からクラスタリング結果に移行する際、あるいは、クラスタリング結果から概要一覧表に戻る際、表示する際の先頭を的確な位置に選ぶことによって、クラスタリング結果と概要一覧表とを相互に表示するような操作を繰り返し行う場合、必要な部分の表示が切れて見えない状態となったり、自分の注目している部分がどこにあるのかがわからなくなることがなくなり、能率よく快適な操作が可能となる。

なお、本実施の形態は上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、本実施形態においては、検索されたそれぞれの文書のタイトルを用いてクラスタリングする例について説明したが、タイトルだけでなく、前述したように、URLアドレス (<http://>を取り除いた部分) を用いたクラスタリングも可能である。

その他、更新日時（単純な時間または最近1カ月以内の更新日時）、ファイルサイズ（web ページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、それぞれのクラスタリング結果に基づいてそれぞれの概要一覧表を作成することができる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。そして、それぞれの検索サービスによって検索された結果に基づいてクラスタリング処理し、そ

れによって得られたクラスタリング結果に基づいて概要一覧表の作成を行うことが可能である。

以上説明したように本実施形態においては、検索された複数の文書をクラスタリング処理し、それによって得られたクラスタリング結果に対し、そのクラスタリング結果の概要が一目でわかるようなクラスタリング結果概要一覧表を作成し、そのクラスタリング概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体のおおよその内容を把握することができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、クラスタリング結果全体のおおよその内容を把握できることから、ユーザは自分の欲しい情報を効率よく探すことができるばかりでなく、思わぬ情報の存在を発見することもでき、新規情報の発掘も容易に行えるようになる。

また、クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクすることで、クラスタリング結果概要一覧表からクラスタリング結果の対応するクラスタ部分に簡単に移行することができ、また、その状態から、クラスタリング結果概要一覧表にも容易に戻ることもできるので、クラスタリング結果が膨大な内容であっても、このような操作を何回か行うことで、やがては欲しい情報の入っているようなクラスタを探すことができ、自分の欲しい情報を効率よく探すことができる。

また、クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果のそれぞれに対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表内のクラスタ名の表示の仕方を変えるようにしたり、ユーザの与えたキーワードを含むクラスタ名の表示の仕方を他のクラスタ名とは異ならせるようにすることで、自分の欲しい情報がどこに入っているかがそのクラスタリング結果概要一覧を見るだけでおおよその見当を付けることができ、自分の欲しい情報を効率よく探すことができる。

請求の範囲

1. 検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力することを特徴とする情報分類方法。

2. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えたことを特徴とする請求項1に記載の情報分類方法。

3. 前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられることを特徴とする請求項2に記載の情報分類方法。

4. 前記複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。

5. 前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。

6. 前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。

7. 前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つであることを特徴とする請求項1から6のいずれか1項に記載の情報分類方法。

8. 請求項1に記載の情報分類方法において、前記クラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理結果のクラスターの順位を再構成し、そのクラスター順位が再構成されたクラスタリング結果を出力することを特徴とする情

報分類方法。

9. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの平均値をそれぞれのクラスタごとに求め、クラスタごとの平均値をそれぞれのクラスタのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

10. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの最大値をそれぞれのクラスタごとに得て、そのクラスタごとのスコアの最大値をそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

11. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書がそれぞれの文書対応に付されたスコアの大きい順に並べられている場合、その中央または中央付近に位置するスコアをそれぞれのクラスタごとに得て、そのクラスタごとの中央または中央付近に位置するスコアをそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

12. 前記クラスタリング処理を複数の検索サービスによって得られた検索結果に対応して行うことを可能とする場合、前記クラスタの順位を再構成するためのクラスタスコアを求める処理は、複数の検索サービスそれぞれに対応して行うことを特徴とする請求項9から11のいずれか1項に記載の情報分類方法。

13. 前記クラスタリング処理は、それぞれの文書のタイトルを検出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいて行うことを特徴とする請求項8から12のいずれか1項に記載の情報分類方法。

14. 前記クラスタ順位が再構成されたクラスタリング結果の出力の仕方は、クラスタスコアの高いクラスタ順に表示し、クラスタスコアが同じであるクラスタが存在する場合には、クラスタ内の文書数の多いクラスタを高順位とすることを特徴とする請求項8から13のいずれか1項に記載の情報分類方法。

15. 請求項1に記載の情報分類方法において、前記クラスタリング処理結果に対し、クラスタリング処理結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング処理結果とともに出力するようにしたことを特徴とする情報分類方法。

16. 前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成されることを特徴とする請求項15に記載の情報分類方法。

17. 前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とすることを特徴とする請求項16に記載の情報分類方法。

18. 前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより1つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行うことを特徴とする請求項17に記載の情報分類方法。

19. 前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行うことを特徴とする請求項18に記載の情報分類方法。

20. 前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うことを特徴とする請求項16から19のいずれか1項に記載の情報分類方法。

21. 前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えることを特徴とする請求項16から20のいずれか1項に記載の情報分類方法。

22. クラスタリング対象となる複数の文書が、ユーザの与えたキーワード

によって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせることを特徴とする請求項16から21のいずれか1項に記載の情報分類方法。

23. 検索サービスで検索された複数の検索結果が入力され、入力された複数の検索結果に対してクラスタリング処理を施し、そのクラスタリング処理結果を出力するクラスタリングモジュールを含んでなることを特徴とする情報分類装置。

24. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールをさらに備えたことを特徴とする請求項23に記載の情報分類装置。

25. 前記クラスタリングモジュールによって得られたクラスタリング結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力するクラスタ順位設定モジュールと、

を有することを特徴とする請求項23に記載の情報分類装置。

26. 前記クラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールと、

を有することを特徴とする請求項23に記載の情報分類装置。

27. 検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を出力する情報分類プログラムを記録した記録媒体であって、

その情報分類プログラムは、前記検索サービスからの検索結果を取得する手順と、

取得した検索結果に対してクラスタリング処理を施し、そのクラスタリング処

理結果を出力する手順とを含んでなることを特徴とする情報分類処理プログラムを記録した記録媒体。

28. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行った後に、前記クラスタリング処理を行うことを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

29. 前記クラスタリング処理によって得られたクラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力する手順と、

を含んでなることを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

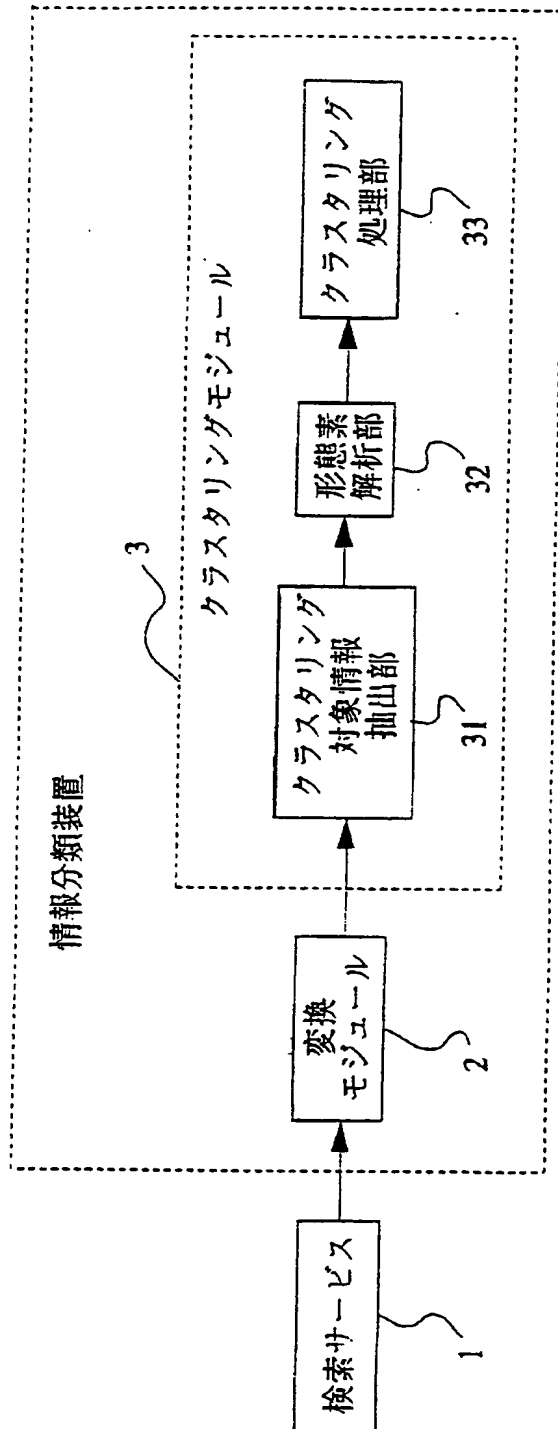
30. 前記クラスタリング処理によって得られたクラスタリング処理結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成手順と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順と、

を含んでなることを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

図 1

1/17



2/17

図 2

- T 1 *用紙カセットについて
A 1 標準装備のユニバーサル用紙カセットはオプションのA 4 専用の大容量用紙カセットに取り替えることが可能である。標準ではトレイに 200 枚の用紙をセットすることが・・・ D 1
- T 2 *レーザプリンタのメモリの増設について
A 2 レーザプリンタのメモリを増設することでパソコンの開放時間を早めたり、・・・することが可能になる。ただし、どの程度・・・ D 2
- T 3 *オプションインターフェースカードの増設について
A 3 オプションのインターフェースカードを使用して、ネットワーク上にプリンタをダイレクトに接続して使用することができる。そして、・・・ D 3
- T 4 *用紙設定で「トレイ」「カセット」「ジドウ」の切換について
A 4 様々なアプリケーションから印刷する際、給紙装置や用紙サイズの設定をする必要がある。用紙カセットには用紙ガイドクリップが・・・ D 4
- T 5 *プリンタにフォントを追加するためにハードディスクを増設する場合について
A 5 フォントを追加する場合は、・・・する方法がある。また、オプションのフォントROMボードを装着する場合は、・・・することができる。 D 5
- T 6 *印刷後における出力された用紙の汚れについて
A 6 用紙の端や裏面に黒い汚れがつく場合には、プリンタ本体定着器ローラのクリーニングを実施する。プリンタ本体のパネルから「クリーニング印刷」の設定とし、A 4 用紙にクリーニング用紙を印刷する。次に・・・ D 6
- T 7 *用紙カセットの増設について
A 7 オプションのダブルカセットユニットもしくはA 4 専用の大容量用紙カセット用ユニットが2つまで装着可能である。・・・ D 7

図 3

3/17.

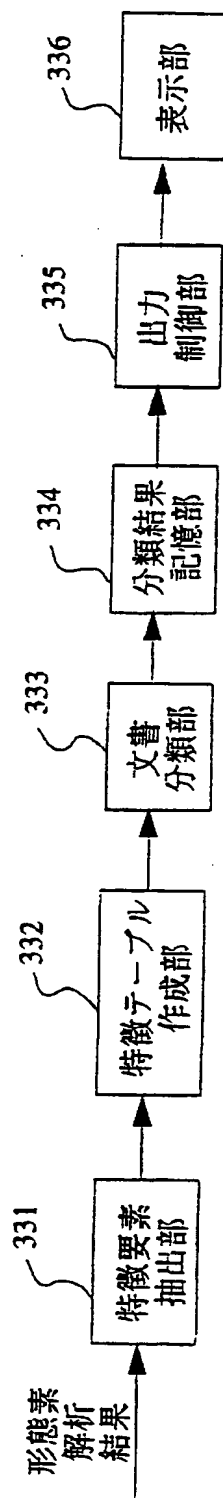


図 4

4/17

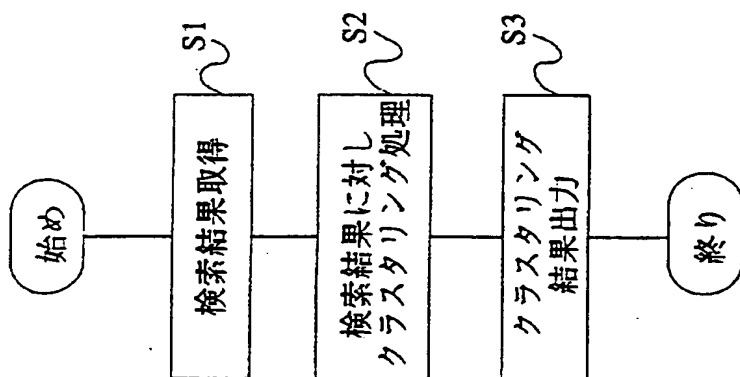


図 5

5/17

特徴要素	文書 D 1	文書 D 2	文書 D 3	文書 D 4	文書 D 5	文書 D 6	文書 D 7
用紙	1			1		1	1
カセット	1			1			1
増設		1	1		1		1

図 6

6/17

特徴要素	クラスタ
用紙	D1, D4, D6, D7
カセット	D1, D4, D7
増設	D2, D3, D5, D7

図 7

7/17

特徴要素	タイトル
用紙	<ul style="list-style-type: none"> * 用紙カセットについて * 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて * 印刷後における出力された用紙の汚れについて * 用紙カセットの増設について
カセット	<ul style="list-style-type: none"> * 用紙カセットについて * 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて * 用紙カセットの増設について
増設	<ul style="list-style-type: none"> * レーザープリンタのメモリ増設について * オプティカルファーストカードの増設について * プリンタにフォントを追加するためにハードディスクを増設する場合について * 用紙カセットの増設について

図 8

8/17

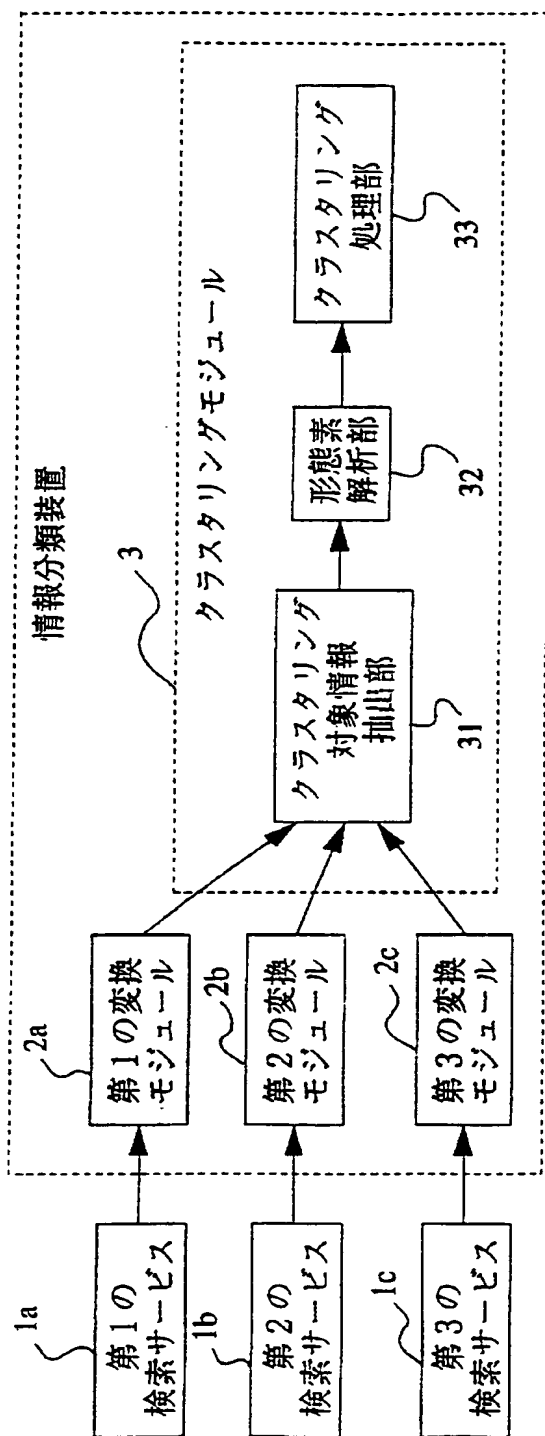


図 9

9/17

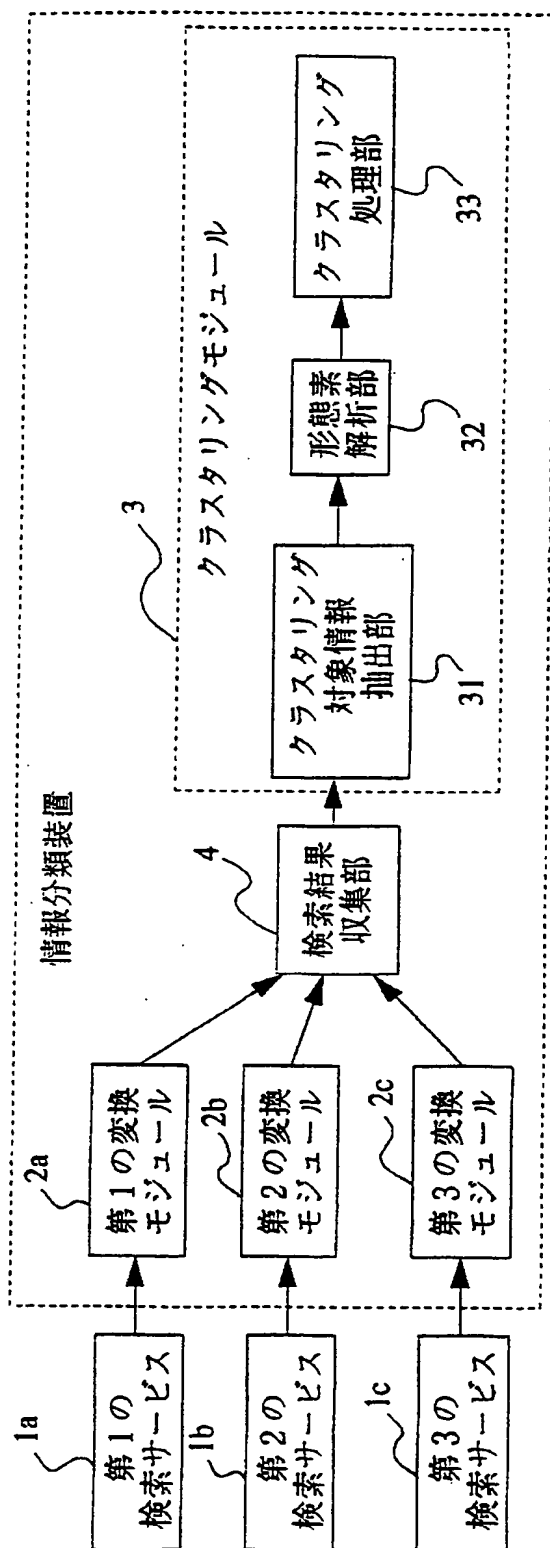
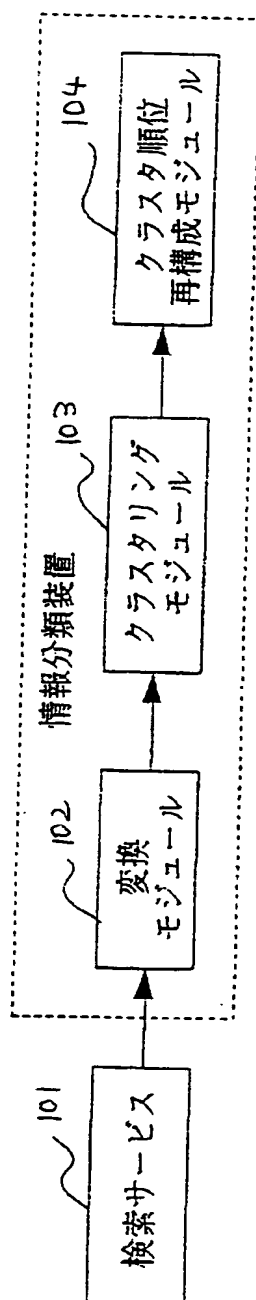


図 10

10/17



11 / 17

図 1 1

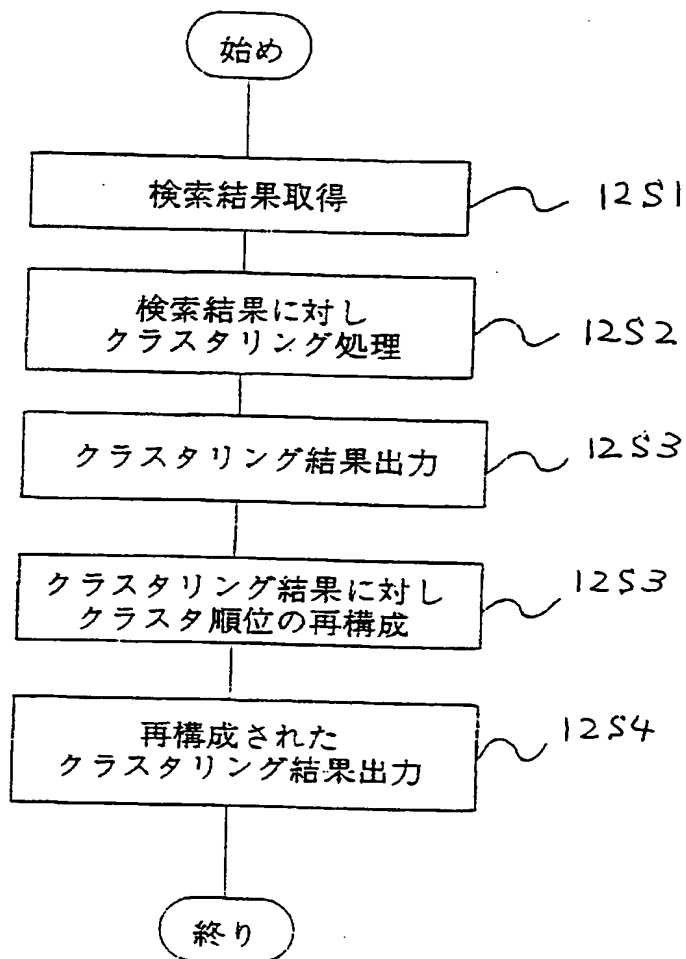
		スコア	文書のタイトル
	文書数	検索結果	
概要	1 6	1 3 3	東北エプソンの概要
		4 9	LP-7000仕様概要
		3 8	LP-8000仕様概要
		3 8	LP-8400仕様概要
		3 4	LP-700仕様概要
		3 3	セイコーエプソン会社概要
		3 0	LP-9200仕様概要
		2 8	LP-9200PS2仕様概要
		2 8	LP-830PS仕様概要
		2 7	LP-8300仕様概要
		2 6	LP-8600仕様概要
		2 4	LP-8200仕様概要
		2 3	LP-710仕様概要
		2 3	LP-800仕様概要
		2 3	LP-900仕様概要
		2 2	LP-500仕様概要
		(クラスタ内合計 579 平均 36)	
LP	1 6	4 9	LP-7000仕様概要
		3 8	LP-8000仕様概要
		3 8	LP-8400仕様概要
		3 4	LP-700仕様概要
		3 0	LP-700S Spec Sheet & Option
		3 0	LP-9200仕様概要
		2 8	LP-9200PS2仕様概要
		2 8	LP-830PS仕様概要
		2 7	LP-8300仕様概要
		2 6	LP-8600仕様概要
		2 4	LP-8200仕様概要
		2 3	LP-710仕様概要
		2 3	LP-800仕様概要
		2 3	LP-900仕様概要
		2 2	LP-500仕様概要
		7	エスパーレーザー LP-8300
		(クラスタ内合計 450 平均 28)	

11/1/17

仕様	14	49 LP-7000仕様概要 38 LP-8000仕様概要 38 LP-8400仕様概要 34 LP-700仕様概要 30 LP-9200仕様概要 28 LP-9200PS2仕様概要 28 LP-830PS仕様概要 27 LP-8300仕様概要 26 LP-8600仕様概要 24 LP-8200仕様概要 23 LP-710仕様概要 23 LP-800仕様概要 23 LP-900仕様概要 22 LP-500仕様概要 (クラスタ内合計 413 平均 29)
デバイス	9	117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP (クラスタ内合計 849 平均 94)
半導体	7	157 半導体事業部環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP (クラスタ内合計 757 平均 108)
電子	4	111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ (クラスタ内合計 349 平均 87)

図 1 2

12/17



13 / 17

図 1 3

クラス名	文書数	検索結果
半導体	7	1 5 7 半導体事業部環境方針 1 1 7 デバイス 半導体 CARD-PC 1 1 5 デバイス 半導体 メモリ 1 1 1 電子デバイス ASIC 1 1 1 デバイス 半導体 PCカード製品 9 0 電子デバイス 半導体 マイコン 5 6 デバイス 半導体 ASSP (クラス内合計 7 5 7 平均 1 0 8)
デバイス	9	1 1 7 デバイス 半導体 CARD-PC 1 1 5 デバイス 半導体 メモリ 1 1 1 電子デバイス ASIC 1 1 1 デバイス 半導体 PCカード製品 1 0 1 デバイス コラボレーション 9 0 電子デバイス 半導体 マイコン 8 2 EPSON 電子デバイス新製品 6 6 EPSON 電子デバイスお問い合わせ 5 6 デバイス 半導体 ASSP (クラス内合計 8 4 9 平均 9 4)
電子	4	1 1 1 電子デバイス ASIC 9 0 電子デバイス 半導体 マイコン 8 2 EPSON 電子デバイス新製品 6 6 EPSON 電子デバイスお問い合わせ (クラス内合計 3 4 9 平均 8 7)

13/4/17

概要	1 6	1 3 3 東北エプソンの概要 4 9 LP-7 0 0 0仕様概要 3 8 LP-8 0 0 0仕様概要 3 8 LP-8 4 0 0仕様概要 3 4 LP-7 0 0 仕様概要 3 3 セイコーエプソン会社概要 3 0 LP-9 2 0 0仕様概要 2 8 LP-9 2 0 0 P S 2仕様概要 2 8 LP-8 3 0 P S仕様概要 2 7 LP-8 3 0 0仕様概要 2 6 LP-8 6 0 0仕様概要 2 4 LP-8 2 0 0仕様概要 2 3 LP-7 1 0仕様概要 2 3 LP-8 0 0仕様概要 2 3 LP-9 0 0仕様概要 2 2 LP-5 0 0仕様概要 (クラスタ内合計 5 7 9 平均 3 6)
仕様	1 4	4 9 LP-7 0 0 0仕様概要 3 8 LP-8 0 0 0仕様概要 3 8 LP-8 4 0 0仕様概要 3 4 LP-7 0 0 仕様概要 3 0 LP-9 2 0 0仕様概要 2 8 LP-9 2 0 0 P S 2仕様概要 2 8 LP-8 3 0 P S仕様概要 2 7 LP-8 3 0 0仕様概要 2 6 LP-8 6 0 0仕様概要 2 4 LP-8 2 0 0仕様概要 2 3 LP-7 1 0仕様概要 2 3 LP-8 0 0仕様概要 2 3 LP-9 0 0仕様概要 2 2 LP-5 0 0仕様概要 (クラスタ内合計 4 1 3 平均 2 9)

13/2/17

LP	16	49 LP-7000仕様概要 38 LP-8000仕様概要 38 LP-8400仕様概要 34 LP-700仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200仕様概要 28 LP-9200PS2仕様概要 28 LP-830PS仕様概要 27 LP-8300仕様概要 26 LP-8600仕様概要 24 LP-8200仕様概要 23 LP-710仕様概要 23 LP-800仕様概要 23 LP-900仕様概要 22 LP-500仕様概要 7 エスパーレーザー LP-8300 (クラスタ内合計 450 平均 28)
----	----	--

図 1 4

14/17

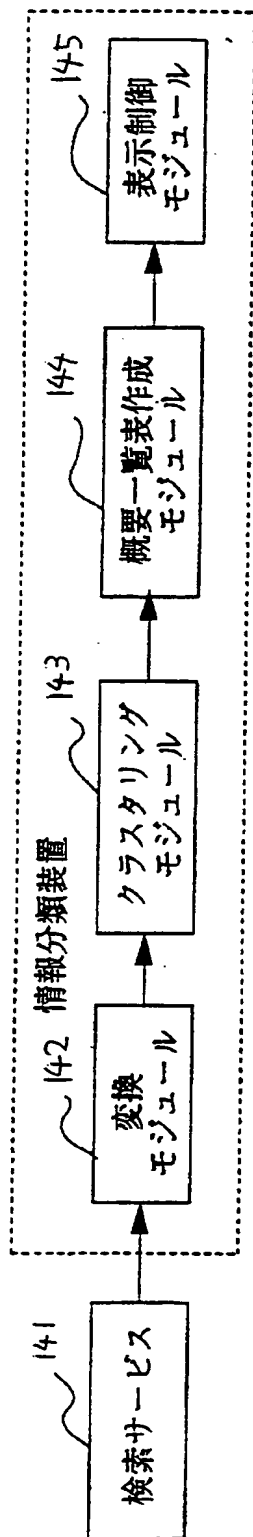
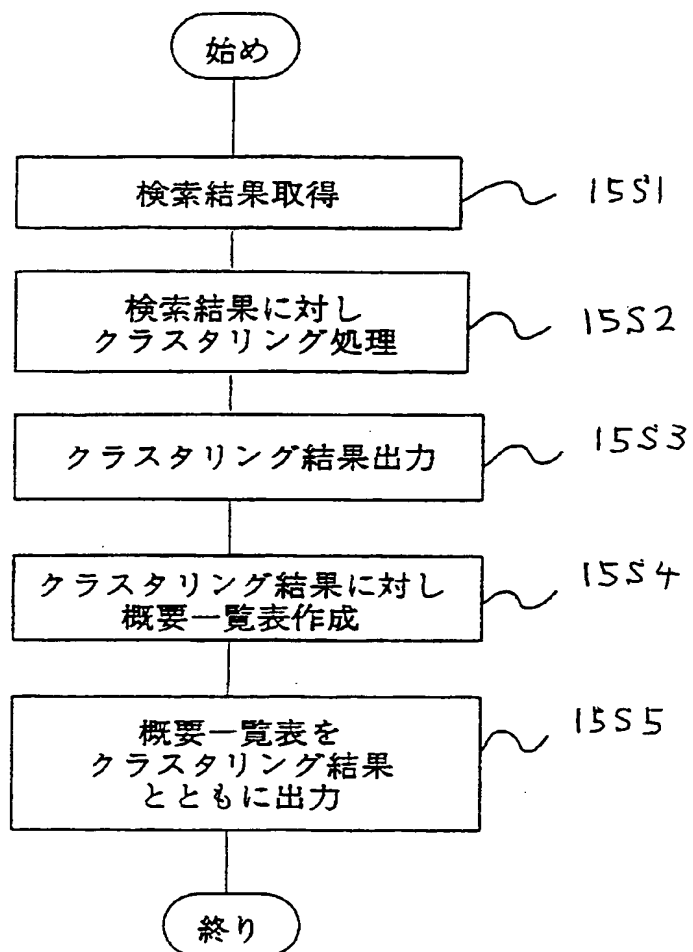


図 1 5

15/17



16 / 17

図 1 6

クラスタリング結果概要一覧					
概要	LP	仕様	デバイス	半導体	電子

1 6 1 0

クラスタ名	文書数	検索結果	
		スコア	文書のタイトル
概要	1 6	1 3 3	東北エプソンの概要
		4 9	LP-7000仕様概要
		3 8	LP-8000仕様概要
		3 8	LP-8400仕様概要
		3 4	LP-700仕様概要
		3 3	セイコーエプソン会社概要
		3 0	LP-9200仕様概要
		2 8	LP-9200PS2仕様概要
		2 8	LP-830PS仕様概要
		2 7	LP-8300仕様概要
		2 6	LP-8600仕様概要
		2 4	LP-8200仕様概要
		2 3	LP-710仕様概要
		2 3	LP-800仕様概要
		2 3	LP-900仕様概要
		2 2	LP-500仕様概要

1 6 2 0

16/4/17

LP	16	49 LP-7000仕様概要 38 LP-8000仕様概要 38 LP-8400仕様概要 34 LP-700仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200仕様概要 28 LP-9200PS2仕様概要 28 LP-830PS仕様概要 27 LP-8300仕様概要 26 LP-8600仕様概要 24 LP-8200仕様概要 23 LP-710仕様概要 23 LP-800仕様概要 23 LP-900仕様概要 22 LP-500仕様概要 7 エスパーレーザー LP-8300
仕様	14	49 LP-7000仕様概要 38 LP-8000仕様概要 38 LP-8400仕様概要 34 LP-700仕様概要 30 LP-9200仕様概要 28 LP-9200PS2仕様概要 28 LP-830PS仕様概要 27 LP-8300仕様概要 26 LP-8600仕様概要 24 LP-8200仕様概要 23 LP-710仕様概要 23 LP-800仕様概要 23 LP-900仕様概要 22 LP-500仕様概要

16/2/17

デバイス	9	1 1 7 デバイス 半導体 CARD-PC 1 1 5 デバイス 半導体 メモリ 1 1 1 電子デバイス ASIC 1 1 1 デバイス 半導体 PCカード製品 1 0 1 デバイス コラボレーション 9 0 電子デバイス 半導体 マイコン 8 2 EPSON 電子デバイス新製品 6 6 EPSON 電子デバイスお問い合わせ 5 6 デバイス 半導体 ASSP
半導体	7	1 5 7 半導体事業部環境方針 1 1 7 デバイス 半導体 CARD-PC 1 1 5 デバイス 半導体 メモリ 1 1 1 電子デバイス ASIC 1 1 1 デバイス 半導体 PCカード製品 9 0 電子デバイス 半導体 マイコン 5 6 デバイス 半導体 ASSP
電子	4	1 1 1 電子デバイス ASIC 9 0 電子デバイス 半導体 マイコン 8 2 EPSON 電子デバイス新製品 6 6 EPSON 電子デバイスお問い合わせ

17 / 17

図 1 7

クラスタリング結果概要一覧		
www.epson.co.jp	www.i-love-epson.co.jp	その他

1 7 4 0

1 7 3 0

クラスタ名	文書数	文書のタイトル 検索結果
www.epson.co.jp	3 1	半導体事業部 環境方針 先輩からのメッセージ - テキスト http://www.epson.co.jp/epson/ae/saiyou/career/... http://www.epson.co.jp/epson/ae/saiyou/career/... 東北エプソンの概要 「vividly NOTE」 http://www.epson.co.jp/epson/ae/saiyou/career/f_... 東北エプソン 先輩からのメッセージ Lattice 社とセイコーエプソン、半導体ビジネス 東北エプソン 企業データ デバイスー半導体ーCARD-PC 事業所別環境方針 デバイスー半導体ーメモリ http://www.epson.co.jp/epson/ae/saiyou/new/a9_... http://www.epson.co.jp/epson/ae/saiyou/new/c_... 職務内容 電子デバイスー半導体ーASIC デバイスー半導体ーPCカード製品 http://www.epson.co.jp/epson/japanese/seihin/de... 東北エプソン 基本ポリシー 酒田5号棟竣工 デバイス コラボレーション ページ EPSON ELECTRONIC DEVICES INDEX 東北エプソン株式会社 環境方針 ムッシュ 上海に新会社 Shanghai Epson Electronics を設... EPSON Ecology 電子デバイスー半導体ーマイクロコンピュータ SEIKO EPSON GROUP http://www.epson.co.jp/epson/japanese/seihin/de... EPSON 電子デバイス新製品

17/1/17

www.i- love- epson.co.jp	1 4	LP-7000仕様概要 LP-8000仕様概要 LP-8400仕様概要 LP-700仕様概要 LP-9200仕様概要 LP-9200PS2仕様概要 LP-830PS仕様概要 LP-8300仕様概要 LP-8600仕様概要 LP-8200仕様概要 LP-710仕様概要 LP-800仕様概要 LP-900仕様概要 LP-500仕様概要
その他		自然環境問題：オゾン層破壊 オゾン層とは KX-Server：PP-430

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl⁷ G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1926-1996 Jitsuyo Shinan Toroku Koho 1996-2000
Kokai Jitsuyo Shinan Koho 1971-2000 Toroku Jitsuyo Shinan Koho 1994-2000

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE ((classification+cluster)*(metasearch)) (in Japanese)
WPI (cluster*metasearch)
INSPEC (cluster*metasearch)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Oren Zamir, Oren Etzioni, "Grouper: a dynamic clustering interface to Web search results," Computer Networks, Vol.31, No.11-16, pp.1361-1374 17 May, 1999 (17.05.99) page 1363, right column, 3 rd line from the bottom to page 1369, right column, line 7	1-14, 23-25, 27-29
Y	page 1363, right column, 3 rd line from the bottom to page 1369, right column, line 7	15-22, 26, 30
Y	JP, 11-149479, A (NEC Corporation), 02 June, 1999 (02.06.99), Full text, (especially, Fig. 3) (Family: none)	15-22, 26, 30
A	Daniel Dreilinger, Adele E. Howe, "Experiences with Selecting Search Engines Using Metasearch," ACM Transaction on Information Systems, Vol.15, No.3, pp.195-222 July 1997 (07.97) Full text	2-6, 24, 28

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search
17 August, 2000 (17.08.00)

Date of mailing of the international search report
29 August, 2000 (29.08.00)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Iwadera, et al., "Tagengo Bunsan Joho Kensaku Architecture ni kansuru Kentou", Research Report of Information Processing Society of Japan (IPSJ), Vol.98, No.82 (98-NL-127), pp.63-70 (Japan) 17 September, 1998 (17.09.98) Full text	2-6, 24,28
A	JP, 7-319905, A (Fujitsu Limited), 08 December, 1995 (08.12.95), Full text (Family: none)	8-14, 25,29
A	JP, 10-74206, A (Matsushita Electric Ind. Co., Ltd.), 17 March, 1998 (17.03.98), Full text, (especially, Par. Nos. [0042] to [0050]; Fig. 9) (Family: none)	8-14, 25,29
A	Shimamura et al., "WWW Document Kensaku ni okeru Domain Mei Clustering no Riyou", Research Report of Information Processing Society of Japan (IPSJ), Vol.98, No.22 (98-HI-77), pp.7-12 (Japan) 13 March 1998 (13.03.98) Full text	15-22,26,30
P,X	Kawamae, et al., "Meta Kensaku System Kouchiku no tame no Data no Kouzouka ni kansuru Kenkyu", Workshop materials of the 38 th Jinko Chino Kiso Kenkyukai and the 45 th Chishiki Base System Kenkyukai pp.37-42 (SIG-FAI/KBS-9902-7), (Japan) 29 September, 1999 (29.09.99) Full text	1-7, 23,24,27,28

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The idea such that "a clustering module obtains the results of search by a search service, clusters the results of search, and outputs the results of the clustering" does not go beyond the prior art and is not a special technical feature provided for in the second paragraph of PCT Rule 13.2. Therefore there are no common subject matters common to the inventions of all the claims.

The subject matters of the inventions of claims 1-7, 23, 24, 27, 28 relate to clustering of the results of metasearch. The subject matters of the inventions of claims 8-14, 25, 29 relate to the order of clusters. The subject matters of the inventions of claims 15-22, 26, 30 relate to a list of the outline of the results of clustering.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

国際調査報告		国際出願番号 PCT/JPO0/03623	
A. 発明の属する分野の分類 (国際特許分類 (IPC))			
Int. Cl ⁷ G06F17/30			
B. 調査を行った分野			
調査を行った最小限資料 (国際特許分類 (IPC))			
Int. Cl ⁷ G06F17/30			
最小限資料以外の資料で調査を行った分野に含まれるもの			
日本国実用新案公報 1926-1996年 日本国公開実用新案公報 1971-2000年 日本国実用新案登録公報 1996-2000年 日本国登録実用新案公報 1994-2000年			
国際調査で利用した電子データベース (データベースの名称、調査に利用した用語) JICST 科学技術文献ファイル (分類+クラス+cluster) * (メタ検索+metasearch) WPI (cluster+metasearch) INSPEC (cluster+metasearch)			
C. 関連すると認められる文献			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号	
X	Oren Zamir, Oren Etzioni, "Grouper: a dynamic clustering interface to Web search results," Computer Networks, Vol.31, No.11-16, p.1361-1374 17.May.1999(17.05.99) 第1363頁右欄下から3行目～第1369頁右欄7行	1-14, 23-25, 27-29	
Y	第1363頁右欄下から3行目～第1369頁右欄7行	15-22, 26, 30	
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。			
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的な技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に言及する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願 の日の後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献			
国際調査を完了した日 17.08.00		国際調査報告の発送日 29.08.00	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 平井 誠 5L 9740 電話番号 03-3581-1101 内線 3560	

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P, 11-149479, A (日本電気株式会社) 2. 6月. 1999 (02. 06. 99) 全文 (特に、図3) (ファミリーなし)	15-22, 26, 30
A	Daniel Dreilinger, Adele E. Howe, "Experiences with Selecting Search Engines Using Metasearch," ACM Transactions on Information Systems, Vol. 15, No. 3, p. 195-222 July. 1997 (07. 97) 全文	2-6, 24, 28
A	巖寺、林、菊井、小橋、Leong M, Choi K, 「多言語分散情報検索 アーキテクチャに関する検討」, 情報処理学会研究報告, Vol. 98, No. 82 (98-NL-127), p. 63-70, (日) 17. 9月. 1998 (17. 09. 98) 全文	2-6, 24, 28
A	J P, 7-319905, A (富士通株式会社) 8. 12月. 1995 (08. 12. 95) 全文 (ファミリーなし)	8-14, 25, 29
A	J P, 10-74206, A (松下電器産業株式会社) 17. 3月. 1998 (17. 03. 98) 全文 (特に、段落【0042】-【0050】、図9) (ファミリーなし)	8-14, 25, 29
A	島村、高野, 「WWWドキュメント検索におけるドメイン名クラス タリングの利用」, 情報処理学会研究報告, Vol. 98, No. 22 (98-HI -77), p. 7-12, (日) 13. 3月. 1998 (13. 03. 98) 全文	15-22, 26, 30
P, X	川前、青木、安田, 「メタ検索システム構築の為のデータの構造化 に関する研究」, 第38回人工知能基礎研究会&第45回知識ベースシ ステム研究会資料, p. 37-42 (SIG-FAI/KBS-9902-7), (日) 29. 9月. 1999 (29. 09. 99) 全文	1-7, 23, 24, 27, 28

第Ⅰ欄 請求の範囲の一部の調査ができないときの意見 (第1ページの2の続き)

法第8条第3項 (PCT17条(2)(a)) の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. ☐ 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第Ⅱ欄 発明の単一性が欠如しているときの意見 (第1ページの3の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

「検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力すること」は、先行技術の域を出ないから、PCT規則13.2の第2文の意味において特別な技術事項でない。それ故、請求の範囲全てに共通の事項はない。

請求の範囲1-7, 23, 24, 27, 28は、メタ検索の結果のクラスタリングに関するものである。請求の範囲8-14, 25, 29は、クラスターの順位に関するものである。請求の範囲15-22, 26, 30は、クラスタリング結果の概要一覧表に関するものである。

1. ☒ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
- ☒ 追加調査手数料の納付と共に出願人から異議申立てがなかった。

P C T

国際調査報告

(法 8 条、法施行規則第40、41条)
[P C T 1 8 条、P C T 規則43、44]

出願人又は代理人 の書類記号 F005275W000	今後の手続きについては、国際調査報告の送付通知様式(P C T / I S A / 2 2 0) 及び下記 5 を参照すること。	
国際出願番号 P C T / J P 0 0 / 0 3 6 2 3	国際出願日 (日.月.年) 0 2 . 0 6 . 0 0	優先日 (日.月.年) 0 4 . 0 6 . 9 9
出願人 (氏名又は名称) セイコーエプソン株式会社		

国際調査機関が作成したこの国際調査報告を法施行規則第41条 (P C T 1 8 条) の規定に従い出願人に送付する。
この写しは国際事務局にも送付される。

この国際調査報告は、全部で 4 ページである。

☐ この調査報告に引用された先行技術文献の写しも添付されている。

1. 国際調査報告の基礎

a. 言語は、下記に示す場合を除くほか、この国際出願がされたものに基づき国際調査を行った。

☐ この国際調査機関に提出された国際出願の翻訳文に基づき国際調査を行った。

b. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際調査を行った。

☐ この国際出願に含まれる書面による配列表

☐ この国際出願と共に提出されたフレキシブルディスクによる配列表

☐ 出願後に、この国際調査機関に提出された書面による配列表

☐ 出願後に、この国際調査機関に提出されたフレキシブルディスクによる配列表

☐ 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった。

☐ 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記載した配列が同一である旨の陳述書の提出があった。

2. ☐ 請求の範囲の一部の調査ができない (第 I 欄参照)。

3. ☒ 発明の単一性が欠如している (第 II 欄参照)。

4. 発明の名称は ☒ 出願人が提出したものを承認する。

☐ 次に示すように国際調査機関が作成した。

5. 要約は ☒ 出願人が提出したものを承認する。

☐ 第 III 欄に示されているように、法施行規則第47条 (P C T 規則38.2(b)) の規定により国際調査機関が作成した。出願人は、この国際調査報告の発送の日から 1 カ月以内にこの国際調査機関に意見を提出することができる。

6. 要約書とともに公表される図は、

第 1 図とする。 ☒ 出願人が示したとおりである。

☐ なし

☐ 出願人は図を示さなかった。

☐ 本図は発明の特徴を一層よく表している。

第I欄 請求の範囲の一部の調査ができないときの意見 (第1ページの2の続き)

法第8条第3項 (PCT17条(2)(a)) の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. ☐ 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第II欄 発明の単一性が欠如しているときの意見 (第1ページの3の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

「検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力すること」は、先行技術の域を出ないから、PCT規則13.2の第2文の意味において特別な技術事項でない。それ故、請求の範囲全てに共通の事項はない。

請求の範囲1-7, 23, 24, 27, 28は、メタ検索の結果のクラスタリングに関するものである。請求の範囲8-14, 25, 29は、クラスタの順位に関するものである。請求の範囲15-22, 26, 30は、クラスタリング結果の概要一覧表に関するものである。

1. ☒ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったので、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったので、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
- ☒ 追加調査手数料の納付と共に出願人から異議申立てがなかった。

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報 1926-1996年
 日本国公開実用新案公報 1971-2000年
 日本国実用新案登録公報 1996-2000年
 日本国登録実用新案公報 1994-2000年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICST 科学技術文献ファイル ((分類+クラス+cluster) * (メタ検索+metasearch))

WPI (cluster*metasearch)

INSPEC (cluster*metasearch)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	Oren Zamir, Oren Etzioni, "Grouper: a dynamic clustering interface to Web search results," Computer Networks, Vol. 31, No. 11-16, p. 1361-1374 17. May. 1999 (17. 05. 99) 第1363頁右欄下から3行目～第1369頁右欄7行	1-14, 23-25, 27-29
Y	第1363頁右欄下から3行目～第1369頁右欄7行	15-22, 26, 30

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
 「O」 口頭による開示、使用、展示等に言及する文献
 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
 「&」 同一パテントファミリー文献

国際調査を完了した日

17. 08. 00

国際調査報告の発送日

29.08.00

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号 100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

平井 誠

5L

9740

電話番号 03-3581-1101 内線 3560

C (続き) 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P, 11-149479, A (日本電気株式会社) 2. 6月. 1999 (02. 06. 99) 全文 (特に、図3) (ファミリーなし)	15-22, 26, 30
A	Daniel Dreilinger, Adele E. Howe, "Experiences with Selecting Search Engines Using Metasearch," ACM Transactions on Information Systems, Vol.15, No.3, p.195-222 July.1997(07.97) 全文	2-6, 24, 28
A	巖寺、林、菊井、小橋、Leong M、Choi K, 「多言語分散情報検索 アーキテクチャに関する検討」, 情報処理学会研究報告, Vol.98, No.82 (98-NL-127), p.63-70, (日) 17. 9月. 1998 (17. 09. 98) 全文	2-6, 24, 28
A	J P, 7-319905, A (富士通株式会社) 8. 12月. 1995 (08. 12. 95) 全文 (ファミリーなし)	8-14, 25, 29
A	J P, 10-74206, A (松下電器産業株式会社) 17. 3月. 1998 (17. 03. 98) 全文 (特に、段落【0042】-【0050】、図9) (ファミリーなし)	8-14, 25, 29
A	島村、高野, 「WWWドキュメント検索におけるドメイン名グラス タリングの利用」, 情報処理学会研究報告, Vol.98, No.22 (98-HI -77), p.7-12, (日) 13. 3月. 1998 (13. 03. 98) 全文	15-22, 26, 30
P, X	川前、青木、安田, 「メタ検索システム構築の為のデータの構造化 に関する研究」, 第38回人工知能基礎研究会&第45回知識ベースシ ステム研究会資料, p.37-42 (SIG-FAI/KBS-9902-7), (日) 29. 9月. 1999 (29. 09. 99) 全文	1-7, 23, 24, 27, 28

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

The idea such that "a clustering module obtains the results of search by a search service, clusters the results of search, and outputs the results of the clustering" does not go beyond the prior art and is not a special technical feature provided for in the second paragraph of PCT Rule 13.2. Therefore there are no common subject matters common to the inventions of all the claims.

The subject matters of the inventions of claims 1-7, 23, 24, 27, 28 relate to clustering of the results of metasearch. The subject matters of the inventions of claims 8-14, 25, 29 relate to the order of clusters. The subject matters of the inventions of claims 15-22, 26, 30 relate to a list of the outline of the results of clustering.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☒ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

A. CLASSIFICATION OF SUBJECT MATTER
Int.Cl⁷ G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
Int.Cl⁷ G06F17/30Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Jitsuyo Shinan Koho 1926-1996 Jitsuyo Shinan Toroku Koho 1996-2000
Kokai Jitsuyo Shinan Koho 1971-2000 Toroku Jitsuyo Shinan Koho 1994-2000Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
JICST FILE ((classification+cluster)*(metasearch))(in Japanese)
WPI (cluster*metasearch)
INSPEC (cluster*metasearch)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Oren Zamir, Oren Etzioni, "Grouper: a dynamic clustering interface to Web search results," Computer Networks, Vol.31, No.11-16, pp.1361-1374 17 May, 1999 (17.05.99) page 1363, right column, 3 rd line from the bottom to page 1369, right column, line 7	1-14, 23-25, 27-29
Y	page 1363, right column, 3 rd line from the bottom to page 1369, right column, line 7	15-22, 26, 30
Y	JP, 11-149479, A (NEC Corporation), 02 June, 1999 (02.06.99), Full text, (especially, Fig. 3) (Family: none)	15-22, 26, 30
A	Daniel Dreilinger, Adele E. Howe, "Experiences with Selecting Search Engines Using Metasearch," ACM Transactionson Information Systems, Vol.15, No.3, pp.195-222 July 1997 (07.97) Full text	2-6, 24, 28

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
17 August, 2000 (17.08.00)Date of mailing of the international search report
29 August, 2000 (29.08.00)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP00/03623

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Iwadera, et al., "Tagengo Bunsan Joho Kensaku Architecture ni kansuru Kentou", Research Report of Information Processing Society of Japan (IPSJ), Vol.98, No.82 (98-NL-127), pp.63-70 (Japan) 17 September, 1998 (17.09.98) Full text	2-6, 24, 28
A	JP, 7-319905, A (Fujitsu Limited), 08 December, 1995 (08.12.95), Full text (Family: none)	8-14, 25, 29
A	JP, 10-74206, A (Matsushita Electric Ind. Co., Ltd.), 17 March, 1998 (17.03.98), Full text, (especially, Par. Nos. [0042] to [0050]; Fig. 9) (Family: none)	8-14, 25, 29
A	Shimamura et al., "WWW Document Kensaku ni okeru Domain Mei Clustering no Riyô", Research Report of Information Processing Society of Japan (IPSJ), Vol.98, No.22 (98-HI-77), pp.7-12 (Japan) 13 March 1998 (13.03.98) Full text	15-22, 26, 30
P, X	Kawamae, et al., "Meta Kensaku System Kouchiku no tame no Data no Kouzouka ni kansuru Kenkyu", Workshop materials of the 38 th Jinko Chino Kiso Kenkyukai and the 45 th Chishiki Base System Kenkyukai pp.37-42 (SIG-FAI/KBS-9902-7), (Japan) 29 September, 1999 (29.09.99). Full text	1-7, 23, 24, 27, 28

国際調査報告

(法 8 条、法施行規則第40、41条)
〔PCT 18条、PCT規則43、44〕

出願人又は代理人 の書類記号 F005275W000	今後の手続きについては、国際調査報告の送付通知様式(PCT/ISA/220) 及び下記5を参照すること。	
国際出願番号 PCT/JP00/03623	国際出願日 (日.月.年) 02.06.00	優先日 (日.月.年) 04.06.99
出願人(氏名又は名称) セイコーエプソン株式会社		

国際調査機関が作成したこの国際調査報告を法施行規則第41条(PCT 18条)の規定に従い出願人に送付する。
この写しは国際事務局にも送付される。

この国際調査報告は、全部で 4 ページである。

☐ この調査報告に引用された先行技術文献の写しも添付されている。

1. 国際調査報告の基礎

a. 言語は、下記に示す場合を除くほか、この国際出願がされたものに基づき国際調査を行った。

☐ この国際調査機関に提出された国際出願の翻訳文に基づき国際調査を行った。

b. この国際出願は、ヌクレオチド又はアミノ酸配列を含んでおり、次の配列表に基づき国際調査を行った。

☐ この国際出願に含まれる書面による配列表

☐ この国際出願と共に提出されたフレキシブルディスクによる配列表

☐ 出願後に、この国際調査機関に提出された書面による配列表

☐ 出願後に、この国際調査機関に提出されたフレキシブルディスクによる配列表

☐ 出願後に提出した書面による配列表が出願時における国際出願の開示の範囲を超える事項を含まない旨の陳述書の提出があった。

☐ 書面による配列表に記載した配列とフレキシブルディスクによる配列表に記載した配列が同一である旨の陳述書の提出があった。

2. ☐ 請求の範囲の一部の調査ができない(第I欄参照)。

3. ☒ 発明の単一性が欠如している(第II欄参照)。

4. 発明の名称は ☒ 出願人が提出したものを承認する。

☐ 次に示すように国際調査機関が作成した。

5. 要約は ☒ 出願人が提出したものを承認する。

☐ 第III欄に示されているように、法施行規則第47条(PCT規則38.2(b))の規定により国際調査機関が作成した。出願人は、この国際調査報告の発送の日から1カ月以内にこの国際調査機関に意見を提出することができる。

6. 要約書とともに公表される図は、

第 1 図とする。 ☒ 出願人が示したとおりである。

☐ なし

☐ 出願人は図を示さなかった。

☐ 本図は発明の特徴を一層よく表している。

第Ⅰ欄 請求の範囲の一部の調査ができないときの意見 (第1ページの2の続き)

法第8条第3項(PCT17条(2)(a))の規定により、この国際調査報告は次の理由により請求の範囲の一部について作成しなかった。

1. ☐ 請求の範囲 _____ は、この国際調査機関が調査をすることを要しない対象に係るものである。つまり、
2. ☐ 請求の範囲 _____ は、有意義な国際調査をすることができる程度まで所定の要件を満たしていない国際出願の部分に係るものである。つまり、
3. ☐ 請求の範囲 _____ は、従属請求の範囲であってPCT規則6.4(a)の第2文及び第3文の規定に従って記載されていない。

第Ⅱ欄 発明の単一性が欠如しているときの意見 (第1ページの3の続き)

次に述べるようにこの国際出願に二以上の発明があるとこの国際調査機関は認めた。

「検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力すること」は、先行技術の域を出ないから、PCT規則13.2の第2文の意味において特別な技術事項でない。それ故、請求の範囲全てに共通の事項はない。

請求の範囲1-7, 23, 24, 27, 28は、メタ検索の結果のクラスタリングに関するものである。請求の範囲8-14, 25, 29は、クラスタの順位に関するものである。請求の範囲15-22, 26, 30は、クラスタリング結果の概要一覧表に関するものである。

1. ☒ 出願人が必要な追加調査手数料をすべて期間内に納付したので、この国際調査報告は、すべての調査可能な請求の範囲について作成した。
2. ☐ 追加調査手数料を要求するまでもなく、すべての調査可能な請求の範囲について調査することができたので、追加調査手数料の納付を求めなかった。
3. ☐ 出願人が必要な追加調査手数料を一部のみしか期間内に納付しなかったため、この国際調査報告は、手数料の納付のあった次の請求の範囲のみについて作成した。
4. ☐ 出願人が必要な追加調査手数料を期間内に納付しなかったため、この国際調査報告は、請求の範囲の最初に記載されている発明に係る次の請求の範囲について作成した。

追加調査手数料の異議の申立てに関する注意

- ☐ 追加調査手数料の納付と共に出願人から異議申立てがあった。
- ☒ 追加調査手数料の納付と共に出願人から異議申立てがなかった。

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int. Cl⁷ G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1926-1996年
日本国公開実用新案公報	1971-2000年
日本国実用新案登録公報	1996-2000年
日本国登録実用新案公報	1994-2000年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

JICST 科学技術文献ファイル (分類+クラス+cluster) * (メタ検索+metasearch)

WPI (cluster*metasearch)

INSPEC (cluster*metasearch)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
X	Oren Zamir, Oren Etzioni, "Grouper: a dynamic clustering interface to Web search results," Computer Networks, Vol. 31, No. 11-16, p. 1361-1374 17. May. 1999 (17.05.99) 第1363頁右欄下から3行目～第1369頁右欄7行	1-14, 23-25, 27-29
Y	第1363頁右欄下から3行目～第1369頁右欄7行	15-22, 26, 30

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

17.08.00

国際調査報告の発送日

29.08.00

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

平井 誠

5L

9740

電話番号 03-3581-1101 内線 3560

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
Y	J P, 11-149479, A (日本電気株式会社) 2. 6月. 1999 (02. 06. 99) 全文 (特に、図3) (ファミリーなし)	15-22, 26, 30
A	Daniel Dreilinger, Adele E. Howe, "Experiences with Selecting Search Engines Using Metasearch," ACM Transactions on Information Systems, Vol.15, No.3, p.195-222 July. 1997(07.97) 全文	2-6, 24, 28
A	巖寺、林、菊井、小橋、Leong M、Choi K, 「多言語分散情報検索 アーキテクチャに関する検討」, 情報処理学会研究報告, Vol.98, No.82 (98-NL-127), p.63-70, (日) 17. 9月. 1998 (17. 09. 98) 全文	2-6, 24, 28
A	J P, 7-319905, A (富士通株式会社) 8. 12月. 1995 (08. 12. 95) 全文 (ファミリーなし)	8-14, 25, 29
A	J P, 10-74206, A (松下電器産業株式会社) 17. 3月. 1998 (17. 03. 98) 全文 (特に、段落【0042】-【0050】, 図9) (ファミリーなし)	8-14, 25, 29
A	島村、高野, 「WWWドキュメント検索におけるドメイン名クラス タリングの利用」, 情報処理学会研究報告, Vol.98, No.22 (98-HI -77), p.7-12, (日) 13. 3月. 1998 (13. 03. 98) 全文	15-22, 26, 30
P, X	川前、青木、安田, 「メタ検索システム構築の為のデータの構造化 に関する研究」, 第38回人工知能基礎研究会&第45回知識ベースシ ステム研究会資料, p.37-42 (SIG-FAI/KBS-9902-7), (日) 29. 9月. 1999 (29. 09. 99) 全文	1-7, 23, 24, 27, 28

明細書

情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体

技術分野

本発明は汎用の検索サービスで検索された結果に対しクラスタリング処理を施すことで、ユーザに見易い形で提示するようにした情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体に関する。

背景技術

ネットワーク上に存在する膨大な量の情報の中からユーザの所望とする情報を検索する場合、検索サービスの存在は重要である。たとえば、インターネットでweb ページを検索する際、ユーザは、幾つかの検索サービスの中から任意の検索サービスを選び、自分の欲しい情報を得るための検索要求としてのキーワードを入力する。これによって、検索サービス側では、入力されたキーワードに基づいて情報検索を行って、その検索結果をユーザに提示する。

しかし、検索サービスによって検索される情報は膨大な量となることも多く、その中からユーザの本当に欲しい情報を見つけるのは非常に大変である。近年、web ページは増大の一途をたどっているため、検索された多数の情報を如何にユーザにわかりやすく提示するかが大きな課題となっている。

最近では、検索された情報をユーザが見やすい形に加工して提示する手法も開発され実用化されつつある。たとえば、ユーザの入力したキーワードで検索された結果から得られるキーワードを用いて再検索することで、情報の絞り込みを行い、ユーザの所望とするweb ページを見つけやすくする方法がある。つまり、検索によって得られる検索結果の集合を特徴づけるキーワードを抽出して、ユーザの本当に欲しい情報の集合に収束させる方法である。

このように、膨大な情報の中から、似た情報の集合を見つけることをクラスタリングという。情報処理ではこのクラスタリングはよく知られた手法であり、膨

大な文書を分類する場合などに一般に使われている。

しかし、現在、一般のユーザに広く利用されている検索サービス（汎用検索サービスという）が検索した結果をクラスタリングすることは行われてはなく、前述したように、入力されたキーワードに基づいて情報を抽出して、抽出された情報を羅列的にユーザに提示するというのが一般的である。したがって、ユーザは、その羅列された多数の情報から自分の欲しい情報を探すという面倒な作業をせざるを得なかった。

そこで本発明は、汎用の検索サービスで得られた検索結果をクラスタリングすることで、ユーザに対し見易い形での検索情報提示を可能とすることを目的としている。

図面の簡単な説明

図1は、本発明の第1の実施形態を説明する図であり、ある1つの検索サービスで得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

図2は、本発明の第1の実施形態で用いられる、ある検索サービスで検索された検索結果としての複数の文書例を示す図である。

図3は、図1で示されたクラスタリング処理部の構成を示すブロック図である。

図4は、第1の実施形態における文書分類処理の手順を概略的に説明するフローチャートである。

図5は、図2で示されたそれぞれの文書のタイトルから抽出された特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルの内容の示す図である。

図6は、図5の特徴テーブルを基にしてそれぞれの文書を分類した分類結果を示す図である。

図7は、図6の分類結果に基づいてそれぞれの文書のタイトルをクラスタリングした例を示す図である。

図8は、選択されたある1つの検索サービスから得られた検索結果に対してクラスタリングを行う場合の情報分類装置の構成を説明するブロック図である。

図9は、複数の検索サービスから得られた検索結果に対してクラスタリングを可能とする場合の情報分類装置の構成を示すものである。

図10は、本発明の第2の実施形態を説明する構成図である。

図11は、ある検索サービスで検索された検索結果としての複数の文書をクラスタリングした結果の一例を示す図である。

図12は、本発明の第2の実施形態における情報分類処理手順を概略的に説明するフローチャートである。

図13は、図11で示されたクラスタリング結果をクラスタ順位再構成処理した結果を示す図である。

図14は、本発明の第3の実施の形態を説明する構成図である。

図15は、本発明の第3の実施形態における情報分類処理手順を概略的に説明するフローチャートである。

図16は、図11で示されたクラスタリング結果とその概要一覧表を表示した例を示す図である。

図17は、URLアドレスでクラスタリング処理して得られたクラスタリング結果とその概要一覧表を表示した例を示す図である。

発明の開示

前述の目的を達成するために、本発明の情報分類方法は、検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力するようにしている。

この情報分類方法において、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えるようにしている。

そして、前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられる。

このとき、複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行

うことも可能であり、また、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことも可能であり、さらに、前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことも可能である。

また、前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つとしている。

また、本発明の情報分類方法において、前記クラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理結果のクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力することもできる。

さらに、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの平均値をそれぞれのクラスタごとに求め、クラスタごとの平均値をそれぞれのクラスタのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成するものである。

また、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの最大値をそれぞれのクラスタごとに得て、そのクラスタごとのスコアの最大値をそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成してもよい。

また、前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書がそれぞれの文書対応に付されたスコアの大きい順に並べられている場合、その中央または中央付近に位置するスコアをそれぞれのクラスタごとに得て、そのクラスタごとの中央または中央付近に位置するスコアをそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成してもよい。

また、前記クラスタリング処理を複数の検索サービスによって得られた検索結果に対応して行うことを可能とする場合、前記クラスタの順位を再構成するため

のクラスタスコアを求める処理は、複数の検索サービスそれぞれに対応して行ってもよい。

また、前記クラスタリング処理は、それぞれの文書のタイトルを検出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいて行ってもよい。

また、前記クラスタ順位が再構成されたクラスタリング結果の出力の仕方は、クラスタスコアの高いクラスタ順に表示し、クラスタスコアが同じであるクラスタが存在する場合には、クラスタ内の文書数の多いクラスタを高順位とすることができる。

また、本発明の情報分類方法において、前記クラスタリング処理結果に対し、クラスタリング処理結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング処理結果とともに出力してもよい。

ここで、前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成される。

また、前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とする。

さらに、前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより1つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行っている。

また、前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行っている。

また、前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うこととしている。

また、前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えている。

さらに、クラスタリング対象となる複数の文書が、ユーザの与えたキーワードによって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせている。

また、本発明の情報分類装置は、検索サービスで検索された複数の検索結果が入力され、入力された複数の検索結果に対してクラスタリング処理を施し、そのクラスタリング処理結果を出力するクラスタリングモジュールを含むものである。

また、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールをさらに備えることもできる。

また、前記クラスタリングモジュールによって得られたクラスタリング結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力するクラスタ順位設定モジュールとを有することもできる。

また、前記クラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールとを有することもできる。

また、本発明の記録媒体は、検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を出力する情報分類プログラムを記録した記録媒体であって、

その情報分類プログラムは、前記検索サービスからの検索結果を取得する手順と、

取得した検索結果に対してクラスタリング処理を施し、そのクラスタリング処

理結果を出力する手順とを含んでいるものである。

ここで、前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行った後に、前記クラスタリング処理を行うこともできる。

また、前記クラスタリング処理によって得られたクラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力する手順とを含んでなるものである。

また、前記クラスタリング処理によって得られたクラスタリング処理結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成手順と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順とを含むものである。

発明を実施するための最良の形態

以下、本発明の実施の形態について説明する。なお、この実施の形態で説明する内容は、本発明の情報分類方法および情報分類装置についての説明であるとともに、本発明の情報分類処理プログラムを記録した記録媒体における情報分類処理プログラムの具体的な処理内容をも含むものである。

(第1の実施形態)

図1は本発明の第1の実施形態を示すもので、大きく分けると、検索サービス1、変換モジュール2、クラスタリングモジュール3とから構成され、変換モジュール2とクラスタリングモジュール3が情報分類装置に相当する。

検索サービス1はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえばwebページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス1で検索された検索結果はファイル形式で出力され、クラスタリングモジュール3に渡されるが、検索サービス1は複数存在し、それ

その検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール2が設けられる。

クラスタリングモジュール3は、検索サービス1により出力された検索結果ファイル内容（変換モジュール1による変換後のファイル内容）からクラスタリングするために必要な情報（クラスタリング対象情報という）を抽出するクラスタリング対象情報抽出部31、そのクラスタリング対象情報として抽出された情報を形態素解析する形態素解析部32、その形態素解析結果に基づいてクラスタリングを行うクラスタリング処理部33などを有している。

クラスタリング対象情報抽出部31は、変換モジュール部2で変換された検索エンジン1の検索結果から、クラスタリング対象情報を抽出するものであり、このクラスタリング対象情報としては幾つか考えられる（後述する）。この実施の形態では、検索結果として抽出された多数の文書のそれぞれのタイトル（見出し）をクラスタリング対象情報として抽出するものとする。たとえば、検索結果として図2に示すような複数の文書D1, D2, ..., D7が得られたとする。これら文書D1, D2, ..., D7はタイトルT1, T2, ..., T7と、それに対する本文A1, A2, ..., A7を持っているものとする。

クラスタリング対象情報抽出部31は、このような検索結果に対し、それぞれの文書D1, D2, ..., D7を解析し、それぞれの文書のタイトルを検出する。このクラスタリング対象情報抽出部31が行うタイトルの検出は、具体的には次のようにして行う。

まず、第1の方法として、文書構造様式によってタイトルと規定される部分があればその部分をタイトルとする。また、第2の方法として、文書構造様式によって、標準より大きな文字で表示する指定がなされている部分があれば、その部分をタイトルとする。また、第3の方法として、定められた数の文または単語を文書先頭より抽出し、その抽出した部分をタイトルとする。さらには、これら第1、第2、第3の方法を順次行い、第1の方法を行ったとき、タイトルと規定されている部分があればその部分をタイトルとし、タイトルと規定される部分が存在しなければ、第2の方法を行い、標準より大きな文字で表示する指定がなされ

ている部分があれば、その部分をタイトルとし、標準より大きな文字で表示する指定がなされていなければ、第3の方法を行ってタイトルを抽出する。

形態素解析部32は、クラスタリング対象情報抽出部31でそれぞれの文書からタイトルとして抽出された部分を形態素解析する。

クラスタリング処理部33は、図3に示すように、特徴要素抽出部331、特徴テーブル作成部332、文書分類部333、分類結果記憶部334、出力制御部335、表示部336などを有している。特徴要素抽出部331は、形態素解析部32で形態素解析された結果から特徴要素を抽出する。

特徴テーブル作成部332は、特徴要素抽出部331で抽出された特徴要素とそれぞれの文書D1～D7との関係を示す特徴テーブルを作成する。なお、この特徴テーブルの具体的な内容については後述する。

文書分類部333は、上述の特徴テーブルの内容を参照し、文書D1, D2, ..., D7を意味的に共通性のある複数のクラスタに分類する。つまり、文書D1, D2, ..., D7のそれぞれのタイトルT1, T2, ..., T7に存在する特徴要素に基づいて、共通する特徴要素を持つ文書を1つのまとまりとし、そのまとまりを1つのクラスタとする。なお、この文書分類部333は同義特徴辞書(図示せず)を有し、共通する特徴要素を持つ文書を1つのまとまりとする処理を行う際、共通する特徴要素であるか否かの判断を、その同義語辞書を用いて同義語が有るか否かにより行い、同義語が存在する場合にはそれを同じクラスタとするというようなことを行うことも可能である。

分類結果記憶部334は、文書分類部333によって分類された内容を記憶する。出力制御部335は分類結果記憶部334の内容を読み出して分類結果として表示部336に表示させる。

このような構成において、本発明の情報分類処理について説明する。本発明が行う情報分類処理手順は概略的には、図4のフローチャートに示すように、まず、汎用の検索エンジンで検索された検索結果を取得し(ステップS1)、取得した検索結果に対しクラスタリング処理を施し(ステップS2)、そのクラスタリング処理結果を出力する(ステップS3)。以下、具体例を参照しながら詳細に説明する。

ここでは、ユーザの入力したキーワードによって図2で示した文書D1, D2, ..., D7が検索結果として検索サービス1から出力されたとする。この検索結果はファイル形式で出力され、変換モジュール2でクラスタリングモジュール3が処理できる形式に変換されたのち、クラスタリングモジュール3に与えられる。

クラスタリングモジュール3に入力されたこれらの文書D1, D2, ..., D7は、クラスタリング対象情報抽出部31にてタイトルが抽出される。たとえば、文書D1についてはタイトルT1が検出され、文書D2についてはタイトルT2が検出され、文書D3についてはタイトルT3が検出されるというように、それぞれの文書D1, D2, ..., D7のタイトルT1, T2, ..., T7が抽出される。

そして、形態素解析部32によってそれぞれのタイトルT1, T2, ..., T7が形態素解析されたのち、その形態素解析結果がクラスタリング処理部33に与えられる。クラスタリング処理部33では、特徴要素抽出部331によって、形態素解析部32からの形態素解析結果に基づいて、それぞれのタイトルT1, T2, ..., T7に存在する特徴要素が抽出される。

そして、特徴テーブル作成部332により、それぞれの特徴要素とその特徴要素をタイトルに含む文書との関係を示す特徴テーブルが作成される。この特徴テーブルの例を図5に示す。なお、ここでは、文書数が3つ以上取り出される特徴要素とその特徴要素をタイトルに含む文書との関係を示し、特徴テーブル内に示される数値は、その特徴要素が各文書のタイトルに幾つ含まれているかの数を示している。たとえば、「用紙」という特徴要素は、文書D1, D4, D6, D7のタイトルT1, T4, T6, T7に、それぞれ1個ずつ含まれていることを示している。

図5の特徴テーブルからもわかるように、「用紙」という特徴要素をタイトルに含む文書は、文書D1, D4, D6, D7であり、また、「カセット」という特徴要素をタイトルに含む文書は、文書D1, D4, D7であり、さらに、「増設」という特徴要素をタイトルに含む文書は、文書D2, D3, D5, D7である。なお、先に説明した図2において、これら各特徴要素部分にはアンダーライ

ンが施されている。

そして、文書分類部 3 3 3 はこのような特徴テーブルを参照して、それぞれの特徴要素ごとのクラスタリングを行う。その分類結果を図 6 に示す。

このような分類結果は分類結果記憶部 3 3 4 に格納される。図 6 に示される分類結果において、たとえば、「用紙」で分類されたクラスタ（文書 D 1，D 4，D 6，D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 1 は用紙カセットについての内容であり、文書 D 4 は用紙設定についての内容であり、文書 D 6 は印刷された後の用紙の汚れについての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

このように、これらの文書 D 1，D 4，D 6，D 7 はどれも用紙に関する内容であり、1 つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

また、「カセット」で分類されたクラスタ（文書 D 1，D 4，D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 1 は用紙カセットについての内容であり、文書 D 4 は用紙設定についての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

また、このように、これらの文書 D 1，D 4，D 6，D 7 にはどれも用紙をセットすることに関する内容が含まれており、1 つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

また、「増設」で分類されたクラスタ（文書 D 2，D 3，D 5，D 7 が含まれる）について見れば、図 2 の文書内容からもわかるように、文書 D 2 はメモリの増設についての内容であり、文書 D 3 はインタフェースカードの増設についての内容であり、文書 D 5 はハードディスクの増設についての内容であり、文書 D 7 は用紙カセットの増設についての内容である。

このように、これらの文書 D 2，D 3，D 5，D 7 はどれも何かを増設する場合についての内容であり、1 つのクラスタとして分類されて何等問題のないものとなり、その分類結果は適切であるといえる。

このような適切な分類が行える理由としては、それぞれの文書のタイトルから特徴要素を抽出し、その特徴要素に基づいて文書を分類しているからである。つ

まり、文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることが多い。したがって、文書のタイトルに含まれる特徴要素を用いて分類を行うことにより、分類結果が散漫になることが少なく、また、ノイズクラスが生成される率も少なくすることができる。また、各文書のタイトルは、その文書の作成者がその文書の主旨を表す内容を表現していることから、文書の制作者側の視点による分類が得られる。

なお、ユーザに実際に提示されるクラスタリング結果は、たとえば図7に示すような内容であり、それぞれの特徴要素とその特徴要素を有するタイトルの一覧が提示される。そして、ユーザはこのようなクラスタリング結果の一覧を見て、自分の欲しい情報の入っているようなタイトル部分をクリックすれば、そのタイトルに対応する本文が表示されるというような表示処理がなされる。

このように、この実施の形態では、ユーザがある汎用の検索サービスを利用し、その検索サービスに何らかのキーワードを入力することによって、複数の文書D1, D2, ..., D7が検索された場合、これら複数の文書D1, D2, ..., D7のタイトルT1, T2, ..., T7を抽出し、そのタイトルに含まれる特徴要素に基づいて、それぞれの文書D1, D2, ..., D7に対しクラスタリング処理を行う。

これによって、従来では、検索サービスによって検索された検索結果は単に羅列的に提示されるだけであったものが、本実施形態においては、検索された検索結果のタイトル内容に基づいたクラスタリング結果をユーザに提示できる。そのクラスタリング結果（図7参照）は、それぞれのタイトルに含まれる特徴要素ごとにそれぞれのタイトルが分類された内容となり、ユーザが見易いように整理されたものとなる。

そして、その中から見たい情報があれば、タイトル部分をクリックすれば、それぞれのタイトルを有する文書を表示させることができる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。

検索サービスはそれぞれに得意の分野があることも多く、たとえば、ある検索

サービスはスポーツ関係の情報を多数保有し、ある検索サービスは学術関係の情報を多数保有し、また、ある検索サービスは芸能関係の情報を多数保有しているというように、それぞれの得意の分野が存在する場合も多い。これらそれぞれの得意分野については豊富な情報を所有しており、ユーザの所望とする情報が適切に取り出される可能性が高い。したがって、情報検索を行う際は、検索サービスを使い分けることも普通に行われる。このように、複数の検索サービスを用いる場合のクラスタリング処理について以下に説明する。

図8は複数の検索サービスを用いてこれまで説明したクラスタリング処理を行うための情報分類装置の構成を説明する図であり、ここでは、複数の検索サービスとして第1の検索サービス1a、第2の検索サービス1b、第3の検索サービス1cの3つの検索サービスが存在するものとする。

このように複数の検索サービス（ここでは検索サービス1a、1b、1c）を対象とする場合には、それぞれの検索サービスにより検索された検索結果の内容、長さ、検索結果出力順序などがまちまちなので、それぞれの検索サービス1a、1b、1cからのファイルをクラスタリングモジュール3で処理可能な形式に変換する変換モジュール2a、2b、2cを検索サービス1a、1b、1cに対応して用意する。なお、クラスタモジュール3の構成は図1と同じであるので同一部分には同一符号が付されている。

このような構成であれば、ユーザは検索しようとする情報の分野に応じて検索サービスを使い分けることができる。たとえば、第1の検索サービス1aがスポーツ関係の情報検索に向いている検索サービスであるとすれば、スポーツ関係の情報を検索しようとする場合、第1の検索サービス1aを用いて検索を行う。また、第2の検索サービス1bが学術関係の情報検索に向いている検索サービスであるとすれば、学術関係の情報を検索しようとする場合、第2の検索サービス1bを用いて検索を行う。

このように、ユーザは検索しようとする情報によって検索サービスを選択することができ、しかも、それによって検索された結果は、クラスタリングモジュール3によってクラスタリング処理がなされ、ユーザが見やすいように整理された状態で提示できる。なお、このクラスタリング処理については、すでに説明した

のでここではその説明は省略する。

このように複数の検索サービスを選択的に使用できるようにすれば、それぞれの得意分野に応じた検索が可能となるばかりでなく、ある1つの検索サービスが混み合っているような場合には、他の検索サービスに切り換えて検索を行うというような柔軟な検索も可能となる。

さらに、複数の検索サービスによる検索処理を並行に行い、それぞれの検索サービスによる検索結果を1つにまとめて、まとめられた内容についてクラスタリング処理することも可能である。これを図9により簡単に説明する。

図9の構成は図8の構成において、第1～第3の検索サービス1a, 1b, 1cに対応して設けられた変換モジュール2a, 2b, 2cとクラスタリングモジュール3との間に、それぞれの検索結果を収集して1つにまとめる検索結果収集部4を設けた構成となっている。その他は、図8と同じ構成であり、同一部分には同一符号が付されている。

このような構成であれば、複数の検索サービス（ここでは第1～第3の検索サービス1a, 1b, 1c）では、ユーザの入力したキーワードに対して並行して検索処理を行い、それぞれの検索サービス1a, 1b, 1cによって検索されたそれぞれの検索結果を、それぞれの検索サービス1a, 1b, 1cに対応する変換モジュール部2a, 2b, 2cでクラスタリングモジュール3が処理可能な形式に変換し、それぞれの変換後のファイルを検索結果収集部4に与え、それぞれの検索結果を1つにまとめる。そして、そのまとめられた検索結果は、クラスタリングモジュール3に入力され、これまで説明したようなクラスタリング処理を行う。

このように、複数の検索サービスを用いて検索処理を行うことにより、1つの検索サービスでは検索できなかった幅広い情報を取得することも可能である。また、検索範囲が広がるので、網羅的に情報を探すことができ、どのような情報が世の中に存在するのかを見極めるのに非常に有効となる。このようにして得られた情報は、これまで説明したようなクラスタリング処理がなされ、見やすい形でユーザに提示される。

また、この図9に示した第1～第3の検索サービス1a, 1b, 1cを用いて

検索処理を並行して行うような場合、それぞれの検索サービス 1 a, 1 b, 1 c で得られた検索結果（変換モジュール 2 a, 2 b, 2 c 出力）を 1 つにまとめたあと、クラスタリング処理を行うのではなく、それぞれの検索サービス 1 a, 1 b, 1 c で得られた検索結果（変換モジュール 2 a, 2 b, 2 c 出力）それぞれについてクラスタリング処理を行い、それぞれのクラスタ処理結果をユーザに提示するようにしてもよい。

これによって、広範囲にわたって存在する情報を効率よく検索することが可能となり、また、ユーザは第 1 ～第 3 の検索サービス 1 a, 1 b, 1 c によるそれぞれの検索結果のそれぞれのクラスタリング結果を見比べたりすることができ、それぞれの検索サービスの特色などを知ることができる。

なお、本実施形態は、上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、上述の実施の形態では、クラスタリングを行うための情報（クラスタリング対象情報）として、検索されたそれぞれの文書のタイトルを用いた例について説明したが、これは、タイトルだけでなく、たとえば、URL アドレス（http://を取り除いた部分）、更新日時（単純な時間または最近 1 カ月以内の更新日時）、ファイルサイズ（web ページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、これらのどれを用いるかは、最初にメニューなどで選択項目を選ぶことで可能となる。また、選んだ項目が無い場合には、他の項目を代用する。たとえば、タイトルを選んだ場合、web ページにタイトルが無い場合には、URL アドレスを代用する。

また、本実施形態における情報分類処理を行う処理プログラムは、フロッピーディスク、光ディスク、ハードディスクなどの記録媒体に記録しておくことができ、本発明はその記録媒体をも含むものである。また、ネットワークから処理プログラムを得るようにしてもよい。

（第 2 の実施形態）

次に、本発明の第2の実施形態について説明する。

第1の実施形態の説明において述べたように、文書のタイトルから特徴要素を抽出してクラスタリングする手法は、演算量や処理時間の面で優れ、適切なクラスタリングが可能となるが、クラスタリングを行うための情報量は文書全体から見れば少ないので、全てが適切にクラスタリングされるとは限らない。特に、タイトルが文書の内容を適切に表していなかったり、文書内容とは大きくかけ離れた奇抜なタイトルが付けられていたりする場合が考えられる。このような場合には、クラスタリング精度は大きく低下し、良好なクラスタリング結果は得られないことになる。

また、特徴要素を抽出して、その特徴要素に基づいてクラスタリングする手法は、特徴要素の頻度などを調べ、それによって、機械的に文書を分類してクラスタリングするものである。このようなクラスタリングでは、文書の意味を解析しているわけではないので、得られたクラスタ（クラスタリングされることによって得られる文書群の1つの集合）が必ずしも意味的な共通性のある文書の集合となるとは限らない。

情報分類においては、以上のような場合であっても、ユーザの検索要求に沿ったクラスタリング結果を提示できるようにすることが好ましい。

本実施形態においては、汎用の検索サービスで得られた検索結果をクラスタリング処理を施し、かつ、クラスタリングによって得られたクラスタの順位を再構成して表示することで、ユーザの検索要求に沿ったクラスタリング結果を提示できるようにしている。

以下に、この第2の実施形態について詳細に説明する。

図10は、第2の実施形態をの装置構成を示すもので、大きく分けると、検索サービス101、変換モジュール102、クラスタリングモジュール103、クラスタ順位再構成モジュール104とから構成され、変換モジュール102、クラスタリングモジュール103、クラスタ順位再構成モジュール104が情報分類装置に相当し、特に、クラスタ順位再構成モジュール104を設けた点に本実施形態の特徴がある。

検索サービス101はインターネットなどで一般に広く使用されている汎用の

検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえば web ページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス 101 で検索された検索結果はファイル形式で出力され、クラスタリングモジュール 103 に渡されるが、複数の検索サービスが存在する場合、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール 102 が設けられる。

クラスタリングモジュール 103 は、検索サービス 101 により出力された検索結果（変換モジュール 102 による変換後のファイル内容）に対しクラスタリング処理を行うもので、この実施の形態では、それぞれの文書から文書のタイトルを抽出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいてクラスタリング処理を行う。

具体的には、それぞれの文書からタイトルとして抽出された部分を形態素解析し、形態素解析された結果から特徴的な用語を特徴要素として抽出する。その後、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成する。この特徴テーブルは、たとえば、抽出されたそれぞれの特徴要素が各文書のタイトルに幾つ含まれているかを、それぞれの特徴要素とそれぞれの文書と対応づけて示すもので、一例として、それぞれの文書のタイトルから、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」というような特徴要素が抽出されたとすると、これらの特徴要素が、それぞれの文書のそれぞれのタイトルに、それぞれ何個含まれているかを示す内容となっている。

このような特徴テーブルに基づいて、多数の文書を意味的に共通性のある複数のクラスタに分類する。つまり、それぞれの文書のそれぞれのタイトルに存在する特徴要素に基づいて、共通する特徴要素を持つ文書を 1 つのまとまりとし、そのまとまりを 1 つのクラスタとする。

このクラスタリングモジュール 103 から、たとえば、図 11 のようなクラスタリング結果が出力されたとする。この図 11 は、前述したように、クラスタリングされることによって得られた各クラスタの名称（上述の特徴要素に相当し、それをここではクラスタ名と呼んでいる）と、そのクラスタに属するそれぞれの

文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、前述したように、それぞれのタイトルごとにスコアを示す数値などが一覧表形式で示されている。

なお、このスコアは、前述したように、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、ここでは、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるとする。このスコアはキーワードとの適合度を表すものであるので、その単位としては、%や点数など検索サービスによって異なるがこの実施の形態では点数で表すものとする。

そして、このクラスタリングモジュール103によってクラスタリングされた段階のクラスタリング結果は、図11に示されるように、クラスタの配置はそれぞれのクラスタに含まれる文書の数の多い順となっている。前述したように、この場合、上から順に、概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタの順となっている。

クラスタ順位再構成モジュール104は、クラスタリングモジュール103で出力されたクラスタリング結果に基づいて、それぞれのクラスタの表示順位を再構成するもので、その処理内容の詳細については後に説明する。

このような構成において、本発明の第2の実施形態について説明する。

本実施形態において行う情報分類処理手順は概略的には、図12のフローチャートに示すように、まず、汎用の検索サービス101で検索された検索結果を取得し（ステップ12S1）、取得した検索結果に対しクラスタリング処理を施し（ステップ12S2）、そのクラスタリング結果を出力する（ステップ12S3）。そして、そのクラスタリング結果に対し、それぞれのクラスタ順位の再構成を行い（ステップ12S4）、再構成されたクラスタリング結果を出力する（ステップ12S5）。以下、具体例を参照しながら詳細に説明する。

この実施の形態では、クラスタリングモジュール103が行うクラスタリング処理は、検索サービス101で検索された文書に対し、それぞれの文書のタイトルを抽出し、そのタイトルから特徴要素を抽出して、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成して、その特徴テーブルの内容に基づいて、それぞれの文書を意味的に共通性のある複数のクラスタに分類する。

また、この実施の形態では、ユーザが「半導体」というキーワードを検索要求として検索サービス101に入力し、それによって得られた多数の文書がクラスタリングモジュール103によってクラスタリングされ、そのクラスタリング結果が図11に示すような結果であったとする。

このクラスタリングモジュール103からのクラスタリング結果は、クラスタ順位再構成モジュール104に入力され、以下に示すような処理がなされる。

まず、図11で示されるクラスタリング結果における各クラスタ（概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタ）において、それぞれのクラスタに含まれる文書対応に付されたスコアを利用して、そのスコアの値の平均を求める。この場合、それぞれのクラスタごとにスコアの値を足し算し、その足し算して得られた結果をそのクラスタに含まれる文書数で割る単純平均を求める。

たとえば、概要クラスタで考えると、この図11に示す検索結果においては、そのクラスタ内のスコアの合計が579点あって、文書数が16個であるので、平均のスコアは約36点と求められる。また、「LP」クラスタで考えると、そのクラスタ内のスコアの合計が450点であって、文書数が16個であるので、平均のスコアは約28点と求められる。同様に、「仕様」クラスタは、そのクラスタ内のスコアの合計が413点であって、文書数が14個であるので、平均のスコアは約29点と求められ、「デバイス」クラスタは、そのクラスタ内のスコアの合計が849点であって、文書数が9個であるので、平均のスコアは約94点と求められ、「半導体」クラスタは、そのクラスタ内のスコアの合計が757点であって、文書数が7個であるので、平均のスコアは約108点と求められ、「電子」クラスタは、そのクラスタ内のスコアの合計が349点であって、文書数が4個であるので、平均のスコアは約87点と求められる。

以上のようにして計算された平均のスコアを各クラスタのスコア（クラスタスコアと呼ぶ）とする。そして、このクラスタスコアの高い順にクラスタの順位を再構成する。

すなわち、この場合、クラスタスコアの最も高いクラスタは、半導体クラスタの108点であり、第2位はデバイスクラスタの94点であり、第3位は電子ク

ラストの87点であり、以下、概要クラスタ(36点)、仕様クラスタ(29点)、LPクラスタ(28点)といった順序となる。

このようにして、それぞれのクラスタごとにクラスタスコアを計算し、求められたクラスタスコアの高い順にクラスタ順位を再構成する。

この再構成されたクラスタリング結果を一覧表形式で表したものが図13である。図13によれば、表の最上段に半導体クラスタが位置し、2番目にデバイスクラスタ、3番目に電子クラスタ、以下、概要クラスタ、仕様クラスタ、LPクラスタといった順序となる。この図13のクラスタリング結果によれば、ユーザの入力した「半導体」というキーワードに対し、そのキーワードに適合する文書が多く含まれるクラスタが上位に来ていることがわかる。

この図13のクラスタリング結果と図11のクラスタリング結果を比較すると、図11のクラスタリング結果では、ユーザの入力した「半導体」というキーワードに対し、そのキーワードとは直接には関係しないような文書で構成される概要クラスタ、LPクラスタ、仕様クラスタといったクラスタが上位に位置し、キーワードに大きく関係するような文書が含まれると思われる半導体クラスタ、デバイスクラスタ、電子クラスタといったクラスタが下位に位置しているが、図13では、それが逆転し、キーワードに大きく関係するような文書が含まれると思われるクラスタが上位に位置するようになる。

なお、クラスタスコアが同じ値となった場合には、クラスタ内に含まれる文書数の多い方を上位とするなどの措置を講ずる。

ここで、各クラスタ内のスコアの合計及び平均については、図13のように表示しても良いし、また、表示しなくても良い。

以上説明したように、単純にそれぞれのクラスタに含まれる文書数(1つのクラスタにまとめられた文書数)によって順位付けするのではなく、それぞれのクラスタごとにそのクラスタに含まれる文書に付されたスコアに基づいてクラスタの順位を決めることによって、キーワードに適合したクラスタ順位が得られる。

なお、図13に示すようなクラスタリング結果がユーザに表示され、ユーザはこのようなクラスタリング結果の一覧表を見て、自分の欲しい情報の入っている文書のタイトル部分をクリックすれば、そのタイトルに対応する本文が表示

されるというような表示処理がなされる。

以上説明したように、本発明の第2の実施形態では、ユーザの入力したキーワードによって検索された多数の文書に対し、これら多数の文書のタイトルに含まれる特徴要素に基づいてクラスタリング処理し、さらに、そのクラスタリング結果に対して、それぞれのクラスタごとにそのクラスタに含まれる文書のスコアの平均を求める。そして、その平均のスコアをクラスタスコアとし、それぞれのクラスタごとのクラスタスコアに基づいて、クラスタ順位の再構成を行う。つまり、クラスタスコアの大きい順にクラスタの並べ替えを行い、図13に示すようなクラスタリング結果として表示する。

これによって、ユーザの欲しい情報の入っけいそうなクラスタが一覧表の上位に位置した状態で表示されているので、自分の欲しい情報を探しやすくなる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。

検索サービスはそれぞれに得意の分野があることも多く、たとえば、ある検索サービスはスポーツ関係の情報を多数保有し、ある検索サービスは学術関係の情報を多数保有し、また、ある検索サービスは芸能関係の情報を多数保有しているというように、それぞれの得意の分野が存在する場合も多い。これらそれぞれの得意分野については豊富な情報を所有しており、ユーザの所望とする情報が適切に取り出される可能性が高い。したがって、情報検索を行う際は、検索サービスを使い分けることも普通に行われる。

このように、複数の検索サービスを用いてクラスタリング処理する場合には、それぞれの検索サービスにより検索された検索結果の内容、長さ、検索結果出力順序などがまちまちなので、それぞれの検索サービスからのファイルをクラスタリングモジュール103で処理可能な形式に変換する変換モジュール102を複数の検索サービスに対応して用意する。そして、さらに、そのクラスタリング結果におけるクラスタ順位再構成を行う場合には、それぞれのクラスタのクラスタスコアを求める処理をそれぞれの検索サービスに対応して行うようにする。

たとえば、本実施形態のクラスタ順位再構成処理についていえば、検索サービ

スによって幾つかの対策を講じる必要がある。たとえば、スコアの分布の幅が非常に大きい場合（たとえば、スコアを表す数値が最大1000から最小は2など）は、対数を取って計算するなどの措置を講じたり、また、きわめてスコアの値が小さい文書（たとえば、殆どの文書が数百のスコアの値があるのに2や3の値しかない文書）はクラスタリング対象から外すといった措置を講じる。

このように、複数の検索サービスに対応できるようにすることで、ユーザは検索しようとする情報の分野に応じて検索サービスを使い分けることができ、それぞれの得意分野に応じた検索が可能となるばかりでなく、ある1つの検索サービスが混み合っているような場合には、他の検索サービスに切り換えて検索を行うというような柔軟な検索も可能となる。

なお、本発明の第2の実施形態は、上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、これまで説明した実施形態では、それぞれのクラスタのクラスタスコアは、そのクラスタに含まれる文書のスコアの単純平均を用いた例について説明したが、このクラスタスコアとしては、それぞれのクラスタ内に含まれる文書のなかで最大のスコアを有する文書のスコアを用いるようにしてもよく、また、それぞれのクラスタ内に含まれる文書に付されたスコアのなかで中央に位置する文書のスコアを用いるようにしてもよい。

このように、クラスタごとのスコアの最大値を用いることで、クラスタ順位を決めるために足し算したり割り算したりという計算を行う必要がなく、計算量を少なくすることができ、しかも、同じクラスタ内に、極端に低いスコアを持つ文書が少数あっても、その影響をあまり受けないようにすることができる。また、クラスタごとのスコアの中央値を用いる場合も、最大値を用いるのと同様、クラスタ順位を決めるための計算量を少なくすることができ、しかも、中央値の場合、同じクラスタ内に、極端に高いスコアや極端に低いスコアを持つ文書が少数あっても、その影響をあまり受けないようにすることができる。

また、本実施形態では、クラスタリングを行うための情報（クラスタリング対象情報）として、検索されたそれぞれの文書のタイトルを用いた例について説明したが、これは、タイトルだけでなく、たとえば、URLアドレス（http://を取

り除いた部分)、更新日時(単純な時間または最近1カ月以内の更新日時)、ファイルサイズ(web ページ本文のバイトサイズなど)を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、これらのどれを用いるかは、最初にメニューなどで選択項目を選ぶことで可能となる。また、選んだ項目が無い場合には、他の項目を代用する。たとえば、タイトルを選んだ場合、web ページにタイトルが無い場合には、URLアドレスを代用する。

(第3の実施形態)

次に、本発明の第3の実施形態について説明する。

情報分類処理において、クラスタリングによって得られたクラスタの数がそれほど多くはない場合、ユーザはそのクラスタリング結果全体を把握するのにそれほど多くの時間を費やさずに済む。

しかし、クラスタリングによって得られたクラスタの数は時として何十個あるいは何百個といった膨大な数量となる場合もある。このような場合、クラスタリング結果全体を見るだけでも大変である。

そこで本発明の第3の実施形態においては、汎用の検索サービスで得られた検索結果をクラスタリング処理を施し、かつ、クラスタリングによって得られたクラスタリング結果全体の概要を一目で把握できるような一覧表を作成することで、ユーザが自分の欲しい情報を効率よく探すことができるようにしている。

以下に、第3の実施形態について詳細に説明する。

図14は本発明の第3の実施形態の概略の構成を示すもので、大きく分けると、検索サービス141、変換モジュール142、クラスタリングモジュール143、クラスタリング結果概要一覧表作成モジュール(以下、概要一覧表作成モジュールという)144、表示制御モジュール145とから構成され、変換モジュール142、クラスタリングモジュール143、概要一覧表作成モジュール144、表示制御モジュール145が情報分類装置に相当し、特に、概要一覧表作成モジ

ジュール 1 4 4 を設けたことに本発明の第 3 の実施形態の特徴がある。

検索サービス 1 4 1 はインターネットなどで一般に広く使用されている汎用の検索サービスであり、ユーザからの検索要求としてのキーワードが入力されることにより、たとえば web ページなどから、入力されたキーワードに基づいた情報検索を行う。この検索サービス 1 4 1 で検索された検索結果はファイル形式で出力され、クラスタリングモジュール 1 4 3 に渡されるが、複数の検索サービスが存在する場合、それぞれの検索サービスによって出力されるデータ形式が異なるので、どのような検索サービスからのファイルであってもその内容を読めるような形式にファイルを変換するために変換モジュール 1 4 2 が設けられる。

クラスタリングモジュール 1 4 3 は、検索サービス 1 4 1 により出力された検索結果（変換モジュール 1 4 2 による変換後のファイル内容）に対しクラスタリング処理を行うもので、この実施の形態では、それぞれの文書から文書のタイトルを抽出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいてクラスタリング処理を行う。

具体的には、それぞれの文書からタイトルとして抽出された部分を形態素解析し、形態素解析された結果から特徴的な用語を特徴要素として抽出する。その後、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成する。この特徴テーブルは、たとえば、抽出されたそれぞれの特徴要素が各文書のタイトルに幾つ含まれているかを、それぞれの特徴要素とそれぞれの文書と対応づけて示すもので、一例として、それぞれの文書のタイトルから、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」というような特徴要素が抽出されたとすると、これらの特徴要素が、それぞれの文書のそれぞれのタイトルに、それぞれ何個含まれているかを示す内容となっている。

このような特徴テーブルに基づいて、多数の文書を意味的に共通性のある複数のクラスタに分類する。つまり、それぞれの文書のそれぞれのタイトルに存在する特徴要素に基づいて、タイトルの中に共通する特徴要素を持つ文書を 1 つのまとまりとし、そのまとまりを 1 つのクラスタとする。

このクラスタリングモジュール 1 4 3 から、たとえば、第 2 の実施形態において説明した図 1 1 のようなクラスタリング結果が出力されたとする。この図 1 1

は前述したように、クラスタリングされることによって得られた各クラスタの名称（上述の特徴要素に相当し、それをここではクラスタ名と呼んでいる）と、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

なお、このスコアは、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、ここでは、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるとする。

そして、このクラスタリングモジュール143によって得られたクラスタリング結果におけるそれぞれのクラスタの配列は、図11に示されるように、ここでは、それぞれのクラスタに含まれる文書の数の多い順となっている。前述したように、この場合、上から順に、概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタの順となっている。

概要一覧表作成モジュール144は、クラスタリングモジュール143から出力されたクラスタリング結果に基づいて、それぞれのクラスタのクラスタ名に基づいて、クラスタリング結果の概要を表すクラスタリング結果概要一覧表（概要一覧表という）を作成するものである。

表示制御モジュール145は、クラスタリングモジュール143によって得られたクラスタリング結果や、概要一覧表作成モジュール144で作成された概要一覧表をユーザに提示するために様々な表示制御を行うもので、この実施の形態では、クラスタリング結果とともに概要一覧表を表示させることは勿論、クラスタリング結果と概要一覧表のリンク部分の相互の表示制御、さらには、注目すべきクラスタをユーザの目に付きやすい表示とするといった様々な表示制御を行う。その具体的な表示制御内容については後述する。

このような構成において、本発明の第3の実施形態の情報分類処理について説明する。本実施形態において行う情報分類処理手順は概略的には、図15のフローチャートに示すように、まず、汎用の検索サービス1で検索された検索結果を取得し（ステップ15S1）、取得した検索結果に対しクラスタリング処理を施し（ステップ15S2）、そのクラスタリング結果を出力する（ステップ15S

3)。そして、そのクラスタリング結果に基づいて概要一覧表作成処理を行い（ステップ15S4）、作成された概要一覧表を上述のクラスタリング結果とともに表示する（ステップ15S5）。なお、この作成された概要一覧表を上述のクラスタリング結果とともに表示する際の表示の仕方としては、たとえば、クラスタリング結果に概要一覧表を画面上で重ねて表示するようにしてもよく、また、概要一覧表とクラスタリング結果を別個に配置し、概要一覧表の後に続いてクラスタリング結果を表示させるようにしてもよい。なお、クラスタリング結果の容量が多い場合には、画面をスクロールすることによって、クラスタリング結果を順次画面上に現すようにする。

以下、本発明の第3の実施形態における情報分類処理内容について具体例を参照しながら詳細に説明する。

この実施の形態では、クラスタリングモジュール143が行うクラスタリング処理は、検索サービス141で検索された文書に対し、それぞれの文書のタイトルを抽出し、そのタイトルから特徴要素を抽出して、抽出された特徴要素とそれぞれの文書との関係を示す特徴テーブルを作成して、その特徴テーブルの内容に基づいて、それぞれの文書を意味的に共通性のある複数のクラスタに分類する。また、この実施の形態では、ユーザが「半導体」というキーワードを検索要求として検索サービス141に入力し、それによって得られた多数の文書がクラスタリングモジュール143によってクラスタリングされ、そのクラスタリング結果が図11に示すような結果であったとする。

このクラスタリングモジュール143からのクラスタリング結果は、概要一覧表作成モジュール144に入力され、以下に示すような処理がなされる。

まず、図11で示されるクラスタリング結果における各クラスタ（概要クラスタ、LPクラスタ、仕様クラスタ、デバイスクラスタ、半導体クラスタ、電子クラスタ）において、それぞれのクラスタ名（「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」）を用いた概要一覧表を作成し、その概要一覧表をクラスタリング結果とともに表示する。

図16は概要一覧表1610をクラスタリング結果1620とともに表示した例を示すもので、この表示例では、概要一覧表1610に続いてクラスタリング

結果1620を表示した例である。ここで用いたクラスタリング結果1620は、クラスタ数が6個ときわめて少ない例であるが、実際には、クラスタ数は何十個あるいは何百個というような場合もあり、その中から自分の欲しい情報を探すには、クラスタリング結果全体を見ながら探すことになるので、単に、クラスタリング結果全体を見せられただけでは自分の欲しい情報を探すのに大変な労力を要する。これに対し、クラスタ名による概要一覧表が表示されることによって、その概要一覧表を構成するクラスタ名を見渡すことで、クラスタリング結果にはどのようなクラスタがあるのか、自分の欲しい情報はその中のどのクラスタに入っているのかといったことをおおよそ見当付けることができる。

そして、概要一覧表1610を構成するそれぞれのクラスタ名とクラスタリング結果とはそれぞれ対応する部分がリンクされている。したがって、たとえば、クラスタリング結果のクラスタ数が多く、一つの画面上ですべてのクラスタリング結果が表示しきれないような場合でも、図11で示される概要一覧表16.1.0の任意のクラスタ名をクリックすると、それに対応するクラスタリング結果1620のクラスタ部分が即座に表示され、その状態で、今度はクラスタリング結果のクラスタ名をクリックすると概要一覧表に即座に戻るといった表示制御が可能となる。

このような表示制御を行う際、使い易さをより一層向上させるために次のような機能を付加する。

まず、概要一覧表におけるクラスタ名の配列順序は、クラスタリング結果におけるそれぞれのクラスタの配列順序に従う。すなわち、図11のクラスタリング結果を例に取れば、そのクラスタリング結果におけるそれぞれのクラスタの配列順序は、それぞれのクラスタに含まれる文書数の多い順となっている。たとえば、概要クラスタには文書数は16個、LPクラスタにも文書数は16個、仕様クラスタには文書数は14個、デバイスクラスタには文書数は9個、半導体クラスタには文書数は7個、電子クラスタには文書数は4個といった具合である。したがって、この場合、概要一覧表においても、図11に示すように、一覧表の左から右に、「概要」、「LP」、「仕様」、「デバイス」、「半導体」、「電子」の順とした配列とする。

なお、クラスタリング結果におけるクラスタ順位の付け方としては、文書数だけではなく、それぞれのクラスタのスコアに基づいて決めることもできる。図 1-1 のクラスタリング結果を見ると、前述したように、そのクラスタに属するそれぞれの文書のタイトルと、そのクラスタには幾つの文書が含まれているかを示す文書数、さらには、それぞれのタイトルごとにスコアを示す数値などが表形式で示されている。

このスコアは、与えられたキーワードに対しその検索サービスが独自の方法で検索したとき、検索された文書対応に付けられた値であり、一般には、与えられたキーワードとそれぞれの文書との適合度を示す客観的な尺度として用いられ、汎用の検索サービスによる検索結果にはこのスコアが付されるのが普通である。

このスコアは、検索方式により計算の仕方や値の考え方などが異なるものの、一般に、値が大きいほど、与えられたキーワードに適合する内容を有する文書であるといえる。

したがって、それぞれのクラスタごとにそれぞれのクラスタにおけるスコアの平均などを求め、その平均のスコアの大きいクラスタ順に配列することも考えられる。このように、クラスタリング結果におけるそれぞれのクラスタ順位を、平均のスコアの大きいクラスタ順とした場合には、それによって作成される概要一覧表のクラスタ名の順序もそれに従う。

たとえば、概要クラスタの平均のスコア（このスコアはキーワードとの適合度を表すものであるので、その単位としては、%や点数など検索サービスによって異なるがこの実施の形態では点数で表すものとする）が約 36 点と求められ、LP クラスタの平均のスコアが約 28 点と求められ、仕様クラスタの平均のスコアが 29 点と求められ、デバイスクラスタの平均のスコアが約 94 点と求められ、半導体クラスタの平均のスコアが 108 点と求められ、電子クラスタの平均のスコアが 87 点と求められたとする。このように計算された平均のスコアを各クラスタのスコア（クラスタスコアと呼ぶ）とする。そして、このクラスタスコアの高い順にクラスタの順位を再構成すると、この場合、クラスタスコアの最も高いクラスタは、半導体クラスタの 108 点であり、第 2 位はデバイスクラスタの 94 点であり、第 3 位は電子クラスタの 87 点であり、以下、概要クラスタ（36 点）、

仕様クラスタ（29点）、LPクラスタ（28点）といった順序となる。

このように、クラスタリング結果におけるそれぞれのクラスタの配列順序が、それぞれのクラスタのクラスタスコアの高い順となっているとすれば、それによって作成される概要一覧表のクラスタ名の順序も、クラスタリング結果のクラスタ配列順序に従って、「半導体」、「デバイス」、「電子」、「概要」、「仕様」、「LP」の順序とする。

ここで、各クラスタ内のスコアの合計及び平均については、図16のように表示しなくても良いし、また、表示しても良い。

一般に、クラスタリング結果のクラスタ順序は何らかの意味を持っているので、概要一覧表のクラスタ名の順序もそれに従うようにした方が、ユーザが情報を探す上で都合がよい場合が多い。ユーザは一般に上から順に見て行く傾向にあるので、概要一覧表のクラスタ名の配置順序をクラスタリング結果におけるクラスタの配置順序と同じにすれば、自分の欲しい情報を探すのに便利である。

次に、概要一覧表1610の或るクラスタ名をクリックしてそれにリンクされたクラスタリング結果1620のクラスタ部分を表示させる際、そのクラスタを囲っている枠（クラスタ枠という）の最上部を先頭にした表示を行う。これは、もしクラスタのクラスタ名部分を基準にしてそのクラスタ名部分を先頭にした表示を行うと、表示させたときに、そのクラスタ名に対応するクラスタ要素（図11ではクラスタに含まれる文書のタイトル）が表示画面上に表示されない場合があるという不都合が生じる。たとえば、画面上にクラスタリング結果1620の半導体クラスタが現れていない状態で、概要一覧表1610の「半導体」というクラスタ名をクリックされたとすると、そのクリック動作とほぼ同時にクラスタリング結果1620の半導体クラスタ部分の表示に移行することになるが、そのとき、そのクラスタ枠内の最上段の行に位置するクラスタ要素（この場合は、「157 半導体事業部環境方針」という文書のタイトル）が表示画面上で見えないということがしばしば生じる。

これを防ぐために、クラスタ枠を基準にしてそのクラスタ枠の最上部を先頭にした表示を行うようにする。これによって、そのクラスタ枠内の最上段に位置する行のクラスタ要素が確実に表示がなされるようになる。

また、さらに確実性を高めるため、それより1つ前に位置するクラスタの最終行のクラスタ要素を先頭にした表示を行うようにしてもよい。たとえば、前述の例で考えれば、概要一覧表1610の「半導体」というクラスタ名をクリックされたとすると、クラスタリング結果の半導体クラスタが表示されることになるが、このとき、それよりも1つ前のデバイスクラスタの最終行のクラスタ要素（図11では、「56 デバイス-半導体-ASSP」）といったクラスタ要素を先頭にした表示を行うようにしてもよい。このように、1つ前に位置するクラスタの最終行のクラスタ要素から表示させるようにすれば、本来表示すべきクラスタのクラスタ要素は確実に表示がなされるようになる。

次に、概要一覧表1610に表示されるそれぞれのクラスタ名は、クラスタリング結果1620におけるそれぞれのクラスタの内容によって表示の大きさや色を変える。このクラスタリング結果におけるそれぞれのクラスタの内容というのは、具体的には、それぞれのクラスタの重要度を示すもので、ユーザの与えたキーワードに対する適合度などであり、それぞれのクラスタに含まれる文書数やそれぞれのクラスタのスコアなどを用いて決める。たとえば、前述したように、それぞれのクラスタにおけるスコアの平均を求め、その平均のスコアの大きいクラスタが重要度の高いクラスタであるとすれば、重要度の高いクラスタに対する概要一覧表1610のクラスタ名の表示の仕方を変える。

たとえば、前述した例によれば、図11のクラスタリング結果におけるそれぞれのクラスタのなかで、半導体クラスタのクラスタスコアが最も高い値であったので、半導体クラスタに対応する概要一覧表のクラスタ名「半導体」の表示の仕方を他のクラスタのクラスタ名と異ならせる。具体的には、「半導体」というクラスタ名を他のクラスタ名と色を変えて表示したり、「半導体」というクラスタ名を囲っている枠だけを他よりも太くしたり、その枠内の面積を他よりも大きな面積としたり、さらにはそのクラスタ名をブリンク表示させたりすることによって、ユーザの目につきやすい表示を行う。

また、それぞれのクラスタにおいてそれぞれのクラスタに含まれる文書数の多いクラスタについても同様に、ユーザの目につきやすい表示を行うようにすることもできる。さらに、スコアの高いクラスタは色を他のクラスタとは異ならせ、

文書数の多いクラスタは枠内の面積を大きくするなど、クラスタの特徴によって表示の仕方を多種多様に設定するようにしてもよい。このようにすれば、どのクラスタがどの同様な特徴を持ったクラスタであるかということが、概要一覧表 1 6 1 0 を見るだけで一目でわかる。

また、概要一覧表 1 6 1 0 においてユーザの入力したキーワードが含まれるクラスタ名は、それをユーザに知らせるために他のクラスタ名と異なった表示を行う。

たとえば、図 1 1 のクラスタリング結果の例では、ユーザの与えたキーワードは「半導体」であるため、クラスタリング結果により得られたクラスタのうち、半導体クラスタはまさにキーワードそのものを含むクラスタである。

この場合、そのクラスタリング結果 1 6 2 0 により作成された概要一覧表 1 6 1 0 において、「半導体」部分は他のクラスタとは異ならせた表示として、たとえば、リンクさせたり、色を変えたり、両者を組み合わせるなど、ユーザの目に付きやすい表示を行う。一般に、ユーザは自分の入力したキーワードと同じものがあるかを探すことが多い。したがって、概要一覧表 1 6 1 0 内に自分の与えたキーワードと同じクラスタ名が有るかどうかを一目でわかるようにすれば、ユーザにとっては自分の欲しい情報を探す上で便利なものとなる。

また、画面上にクラスタリング結果 1 6 2 0 の或る部分が表示されている状態のときに、その表示内容から或るクラスタ名をクリックすると概要一覧表 1 6 1 0 の表示に戻るが、そのとき、概要一覧表 1 6 1 0 の先頭に戻って、概要一覧表 1 6 1 0 の先頭部分からの表示がなされるようにする方が都合がよい。

概要一覧表 1 6 1 0 はクラスタリング結果の概要を簡略化して表すものではあるが、場合によってはそのサイズがきわめて大きなものとなることもある。また、複数の概要一覧表が作成される場合もある。すなわち、これまでの説明では、それぞれの文書のタイトルを用い、そのタイトルによってクラスタリング処理しそれによって得られたクラスタリング結果に基づいて概要一覧表を作成する例について説明したが、クラスタリング処理は、タイトルだけでなく、URL アドレス (<http://>を取り除いた部分) などの情報を用いても行うことができる。

たとえば、図 1 1 のクラスタリング結果を得るために用いた多数の文書に対し、

URLを用いてクラスタリング処理し、それによって得られたクラスタリング結果に基づいて概要一覧表を作成することもできる。

図17は図11と同じ文書に対し、URLアドレスによってクラスタリングされたクラスタリング結果1730とそれに基づいて作成された概要一覧表1740の例を示すものである。この場合、クラスタリングされることによって得られたクラスタのクラスタ名としては、クラスタリングする際の用いられたそれぞれのURLアドレス「www.epson.co.jp」と「www.i-love-epson.ne.jp」や「その他」であり、概要一覧表1740を構成するクラスタ名は、これらのURLアドレス「www.epson.co.jp」と「www.i-love-epson.ne.jp」さらに「その他」が用いられる。

この例で説明したように、クラスタリング処理は、幾つもの方法によってなされる可能性があり、その場合、それぞれによって得られたクラスタリング結果に対する複数の概要一覧表が作成されることになる。

このように、複数の概要一覧表が作成される場合や、一つの概要一覧表であっても一覧表のデータ量がきわめて大きい場合、画面上でクラスタリング結果を見ていて、それに対応する概要一覧表のクラスタ名部分に戻したとき、対応するクラスタ名部分からの表示を行うのではなく、複数の概要一覧表が存在する場合には、複数の概要一覧表のなかで先頭に位置する概要一覧表の先頭部分からの表示がなされるようにする。また、1つの概要一覧表の場合でも、その概要一覧表の先頭部分からの表示がなされるようにする。

これは、たとえば、ユーザがクラスタリング結果を見ていて、もう一度、概要一覧表に戻ってクラスタリング結果全体を見たいというような場合に対処するものである。このような場合、概要一覧表の対応するクラスタ名部分に戻ると、戻った位置が一覧表全体のどの位置かがわからなくなる可能性があるが、概要一覧表が複数存在していても、あるいは概要一覧表のデータ量が多くても、概要一覧表全体の先頭部分からの表示がなされるようにすれば、一覧表全体を見渡すことができる。

以上説明したような種々の機能を追加することによって、概要一覧表の有用性をより一層高めることができる。

以上説明したように、この実施の形態では、膨大な情報をクラスタリングし、そのクラスタリング結果の概要が一目でわかるような概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体を見渡すことができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、その概要一覧表とクラスタリング結果との間に前述したような種々の機能を持たせることによって、ユーザの与えたキーワードに対し重要度の高いクラスタはどれであるかなどが一目でわかり、また、概要一覧表からクラスタリング結果に移行する際、あるいは、クラスタリング結果から概要一覧表に戻る際、表示する際の先頭を的確な位置に選ぶことによって、クラスタリング結果と概要一覧表とを相互に表示するような操作を繰り返し行う場合、必要な部分の表示が切れて見えない状態となったり、自分の注目している部分がどこにあるのかがわからなくなることがなくなり、能率よく快適な操作が可能となる。

なお、本実施の形態は上記の内容に限定されるものではなく、上記の要旨を逸脱しない範囲で種々変形実施可能となるものである。たとえば、本実施形態においては、検索されたそれぞれの文書のタイトルを用いてクラスタリングする例について説明したが、タイトルだけでなく、前述したように、URLアドレス (<http://>を取り除いた部分) を用いたクラスタリングも可能である。

その他、更新日時（単純な時間または最近1カ月以内の更新日時）、ファイルサイズ（web ページ本文のバイトサイズなど）を用いてクラスタリングすることもできる。また、これらは、単独で用いてクラスタリングするようにしてもよく、幾つかを組み合わせてもよい。このように、クラスタリング対象情報を種々選ぶことによって、それぞれに応じた特色のあるクラスタリングが行える。そして、それぞれのクラスタリング結果に基づいてそれぞれの概要一覧表を作成することができる。

また、これまでの説明では、ある1つの汎用の検索サービスで検索された結果をクラスタリング処理する場合について説明したが、複数の検索サービスにより検索された結果をクラスタリング処理する場合にも適用できる。そして、それぞれの検索サービスによって検索された結果に基づいてクラスタリング処理し、そ

れによって得られたクラスタリング結果に基づいて概要一覧表の作成を行うことが可能である。

以上説明したように本実施形態においては、検索された複数の文書をクラスタリング処理し、それによって得られたクラスタリング結果に対し、そのクラスタリング結果の概要が一目でわかるようなクラスタリング結果概要一覧表を作成し、そのクラスタリング概要一覧表をクラスタリング結果とともに表示するようにしたので、クラスタリング結果が膨大な内容となったとしても、クラスタリング結果全体のおおよその内容を把握することができ、ユーザが自分の欲しい情報を探す際の大きな助けとなる。また、クラスタリング結果全体のおおよその内容を把握できることから、ユーザは自分の欲しい情報を効率よく探すことができるばかりでなく、思わぬ情報の存在を発見することもでき、新規情報の発掘も容易に行えるようになる。

また、クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクすることで、クラスタリング結果概要一覧表からクラスタリング結果の対応するクラスタ部分に簡単に移行することができ、また、その状態から、クラスタリング結果概要一覧表にも容易に戻ることもできるので、クラスタリング結果が膨大な内容であっても、このような操作を何回か行うことで、やがては欲しい情報の入っているようなクラスタを探すことができ、自分の欲しい情報を効率よく探すことができる。

また、クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表内のクラスタ名の表示の仕方を変えるようにしたり、ユーザの与えたキーワードを含むクラスタ名の表示の仕方を他のクラスタ名とは異ならせるようにすることで、自分の欲しい情報がどこに入っているかがそのクラスタリング結果概要一覧を見るだけでおおよその見当を付けることができ、自分の欲しい情報を効率よく探すことができる。

請求の範囲

1. 検索サービスで検索された複数の検索結果をクラスタリングモジュールが取得して、そのクラスタリングモジュールでは、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を出力することを特徴とする情報分類方法。
2. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールを備えたことを特徴とする請求項1に記載の情報分類方法。
3. 前記変換モジュールは、クラスタリング処理を複数の検索サービスに対応して行う場合、複数の検索サービスそれぞれに対応して設けられることを特徴とする請求項2に記載の情報分類方法。
4. 前記複数の検索サービスのいずれかを選択的に用いて検索処理を行い、選択された検索サービスで検索された検索結果に対してクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。
5. 前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果を収集して、その収集された検索結果に対してクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。
6. 前記複数の検索サービスの少なくとも2つの検索サービスを用いて並行的な検索処理を行い、それぞれの検索結果に対し個別にクラスタリング処理を行うことを特徴とする請求項3に記載の情報分類方法。
7. 前記検索結果に対しクラスタリング処理を行う際、クラスタリング対象とする情報は、それぞれの検索結果における文書のタイトル、URLアドレス、更新日時、ファイルサイズの少なくとも1つであることを特徴とする請求項1から6のいずれか1項に記載の情報分類方法。
8. 請求項1に記載の情報分類方法において、前記クラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理結果のクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力することを特徴とする情

報分類方法。

9. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの平均値をそれぞれのクラスタごとに求め、クラスタごとの平均値をそれぞれのクラスタのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

10. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書対応に付されたスコアの最大値をそれぞれのクラスタごとに得て、そのクラスタごとのスコアの最大値をそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

11. 前記それぞれのクラスタの順位を再構成する処理は、それぞれのクラスタに含まれるそれぞれの文書がそれぞれの文書対応に付されたスコアの大きい順に並べられている場合、その中央または中央付近に位置するスコアをそれぞれのクラスタごとに得て、そのクラスタごとの中央または中央付近に位置するスコアをそれぞれのクラスタスコアとし、そのクラスタスコアによって、クラスタの順位を再構成することを特徴とする請求項8に記載の情報分類方法。

12. 前記クラスタリング処理を複数の検索サービスによって得られた検索結果に対応して行うことを可能とする場合、前記クラスタの順位を再構成するためのクラスタスコアを求める処理は、複数の検索サービスそれぞれに対応して行うことを特徴とする請求項9から11のいずれか1項に記載の情報分類方法。

13. 前記クラスタリング処理は、それぞれの文書のタイトルを検出し、そのタイトルに含まれる特徴的な用語を特徴要素として抽出し、抽出された特徴要素に基づいて行うことを特徴とする請求項8から12のいずれか1項に記載の情報分類方法。

14. 前記クラスタ順位が再構成されたクラスタリング結果の出力の仕方は、クラスタスコアの高いクラスタ順に表示し、クラスタスコアが同じであるクラスタが存在する場合には、クラスタ内の文書数の多いクラスタを高順位とすることを特徴とする請求項8から13のいずれか1項に記載の情報分類方法。

15. 請求項1に記載の情報分類方法において、前記クラスタリング処理結果に対し、クラスタリング処理結果の概要を表すクラスタリング結果概要一覧表を作成し、そのクラスタリング結果概要一覧表を前記クラスタリング処理結果とともに出力するようにしたことを特徴とする情報分類方法。

16. 前記クラスタリング結果概要一覧表は、クラスタリング処理されることによって得られたそれぞれのクラスタのクラスタ名によって構成されることを特徴とする請求項15に記載の情報分類方法。

17. 前記クラスタリング結果とクラスタリング結果概要一覧表とを相互にリンクし、クラスタリング結果概要一覧表の所定のクラスタ名部分が指示されると、前記クラスタリング結果の対応するクラスタ部分の表示を可能とし、クラスタリング結果の或るクラスタ部分を指示すると、クラスタリング結果概要一覧表の表示を可能とすることを特徴とする請求項16に記載の情報分類方法。

18. 前記クラスタリング結果概要一覧表の或るクラスタ名部分が指示されることによって、前記クラスタリング結果内の対応するクラスタ部分の表示を行う際、そのクラスタを囲む枠の最上部またはそれより1つ手前に存在するクラスタの枠内に存在する最終行部分を先頭にした表示を行うことを特徴とする請求項17に記載の情報分類方法。

19. 前記クラスタリング結果の或るクラスタ部分が指示されることによって、前記クラスタリング結果概要一覧表の表示を行う際、クラスタリング結果概要一覧表の先頭部分からの表示を行うことを特徴とする請求項18に記載の情報分類方法。

20. 前記クラスタリング結果概要一覧表を構成するそれぞれのクラスタ名の配列順序は前記クラスタリング結果内のクラスタ配列順序に従うことを特徴とする請求項16から19のいずれか1項に記載の情報分類方法。

21. 前記クラスタリング結果概要一覧表の表示を行う際、前記クラスタリング結果のそれぞれ対応するクラスタの重要度に応じて、クラスタリング結果概要一覧表上におけるクラスタ名の表示の仕方を変えることを特徴とする請求項16から20のいずれか1項に記載の情報分類方法。

22. クラスタリング対象となる複数の文書が、ユーザの与えたキーワード

によって検索された文書である場合、ユーザの与えたキーワードを含むクラスタ名は前記クラスタリング結果概要一覧表上においてそのクラスタ名の表示の仕方を他のクラスタ名とは異ならせることを特徴とする請求項 16 から 21 のいずれか 1 項に記載の情報分類方法。

23. 検索サービスで検索された複数の検索結果が入力され、入力された複数の検索結果に対してクラスタリング処理を施し、そのクラスタリング処理結果を出力するクラスタリングモジュールを含んでなることを特徴とする情報分類装置。

24. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュールをさらに備えたことを特徴とする請求項 23 に記載の情報分類装置。

25. 前記クラスタリングモジュールによって得られたクラスタリング結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力するクラスタ順位設定モジュールと、

を有することを特徴とする請求項 23 に記載の情報分類装置。

26. 前記クラスタリングモジュールによって得られたクラスタリング結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成部と、

そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する表示制御モジュールと、

を有することを特徴とする請求項 23 に記載の情報分類装置。

27. 検索サービスで検索されたユーザの検索要求に対する複数の検索結果をクラスタリングモジュールがクラスタリング処理してそのクラスタリング処理結果を出力する情報分類プログラムを記録した記録媒体であって、

その情報分類プログラムは、前記検索サービスからの検索結果を取得する手順と、

取得した検索結果に対してクラスタリング処理を施し、そのクラスタリング処

理結果を出力する手順とを含んでなることを特徴とする情報分類処理プログラムを記録した記録媒体。

28. 前記検索サービスで検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する手順を行った後に、前記クラスタリング処理を行うことを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

29. 前記クラスタリング処理によって得られたクラスタリング処理結果に対し、検索されたそれぞれの文書対応に付された検索要求との適合性を示すスコアを用いて、前記クラスタリング処理によって得られたそれぞれのクラスタの順位を再構成し、そのクラスタ順位が再構成されたクラスタリング結果を出力する手順と、

を含んでなることを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

30. 前記クラスタリング処理によって得られたクラスタリング処理結果に対し、クラスタリング結果の概要を表すクラスタリング結果概要一覧表を作成する概要一覧表作成手順と、

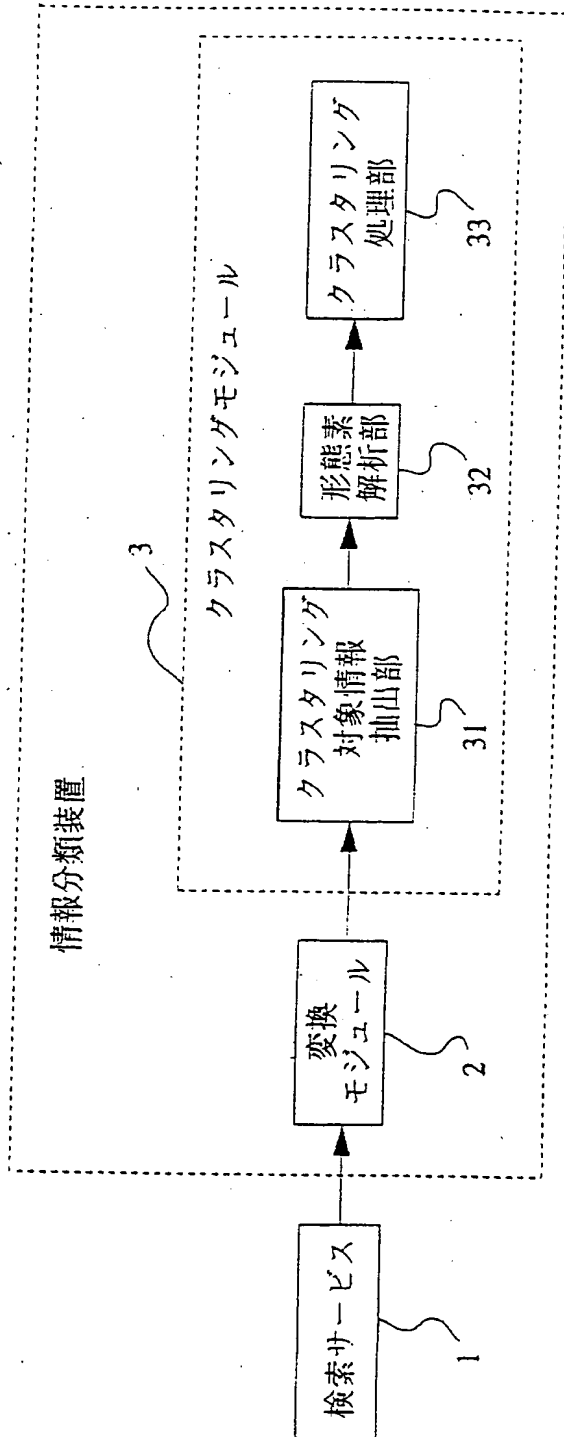
そのクラスタリング結果概要一覧表を前記クラスタリング結果とともに出力する手順と、

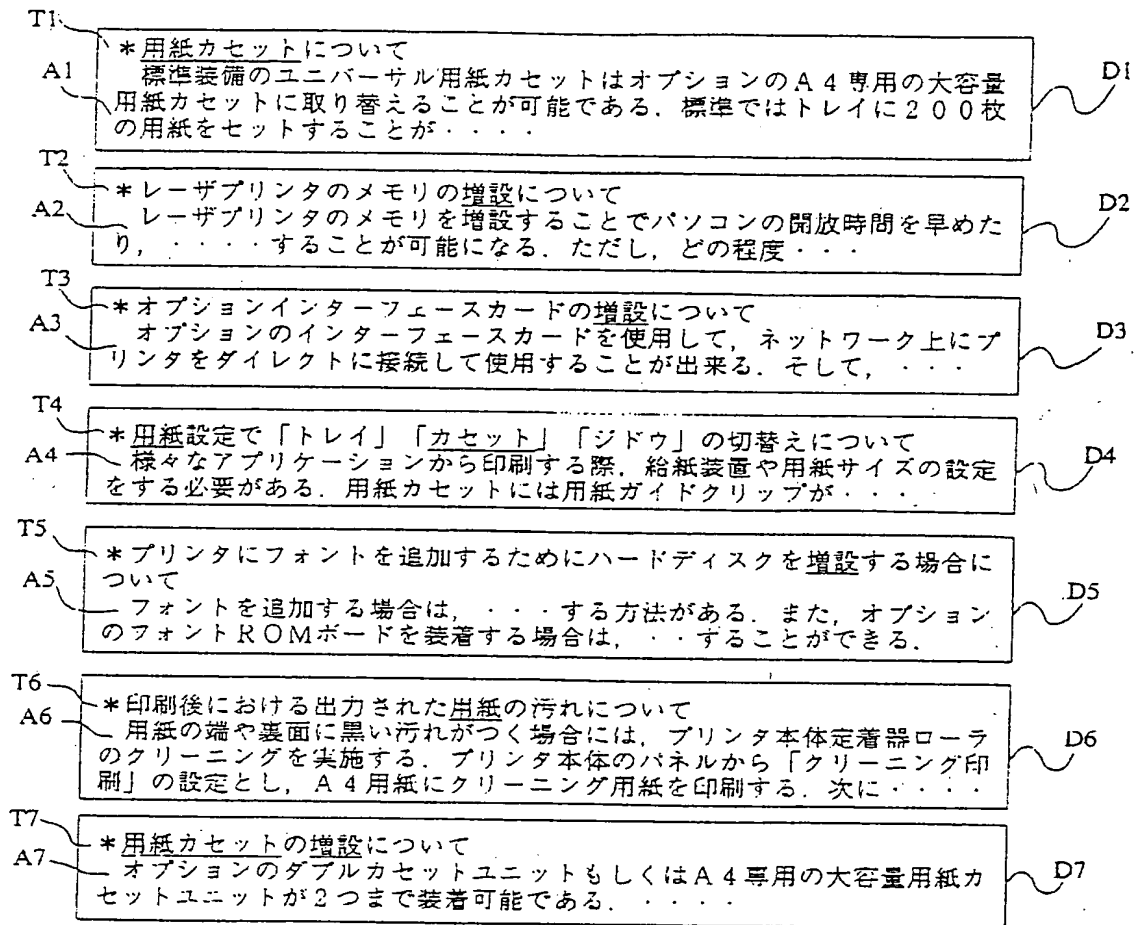
を含んでなることを特徴とする請求項27に記載の情報分類プログラムを記録した記録媒体。

要約書

汎用の検索サービスを用いて検索を行う場合、ユーザの入力するキーワードに対して検索された多数の情報は羅列的に提示されるだけであり、検索結果が多いと検索結果の内容の把握がしにくいという課題があった。

汎用の検索サービス 1 で検索された複数の検索結果をクラスタリングモジュール 3 が取得して、そのクラスタリングモジュール 3 では、その検索結果に対しクラスタリング処理を施し、そのクラスタリング処理結果を提示する。このとき、検索サービス 1 で検索された検索結果を前記クラスタリングモジュールが処理可能な形式に変換する変換モジュール 2 を備える。





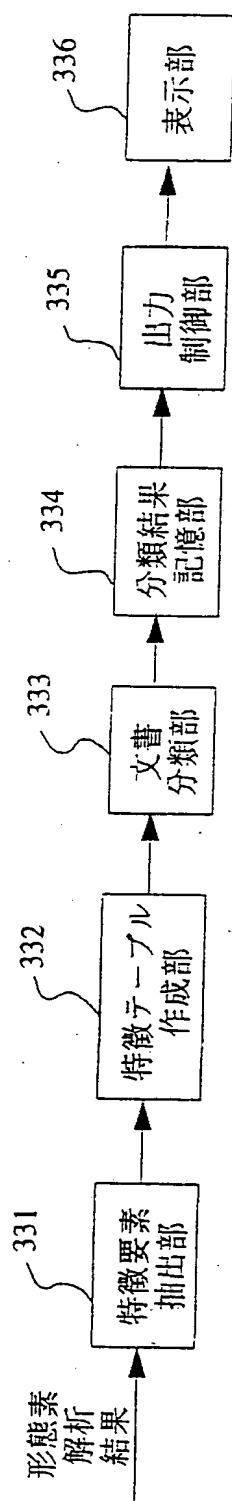
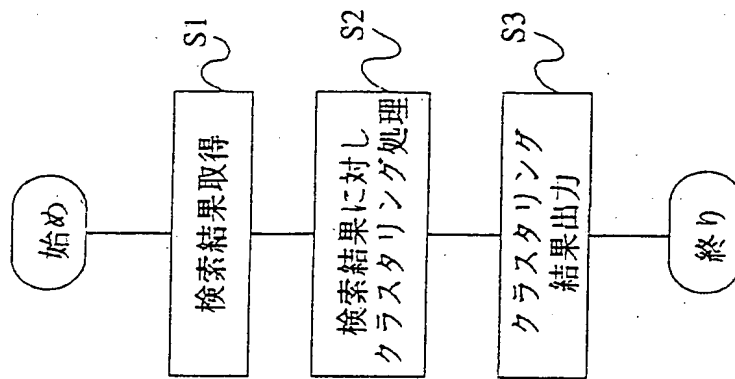


図 4

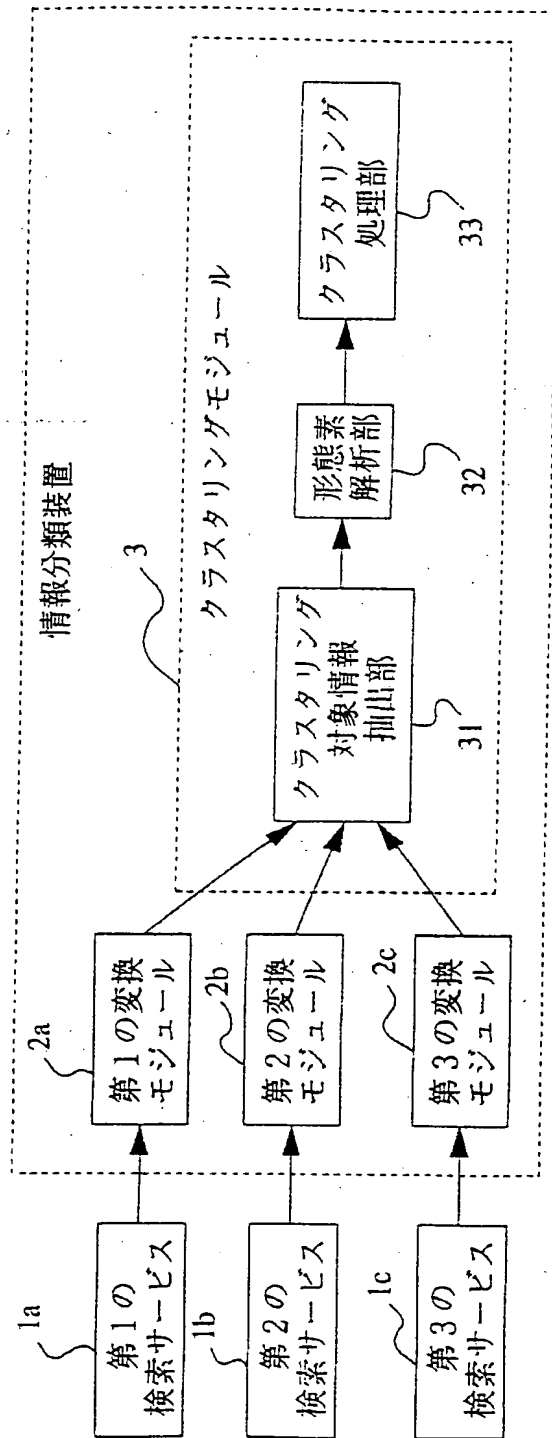
4/17

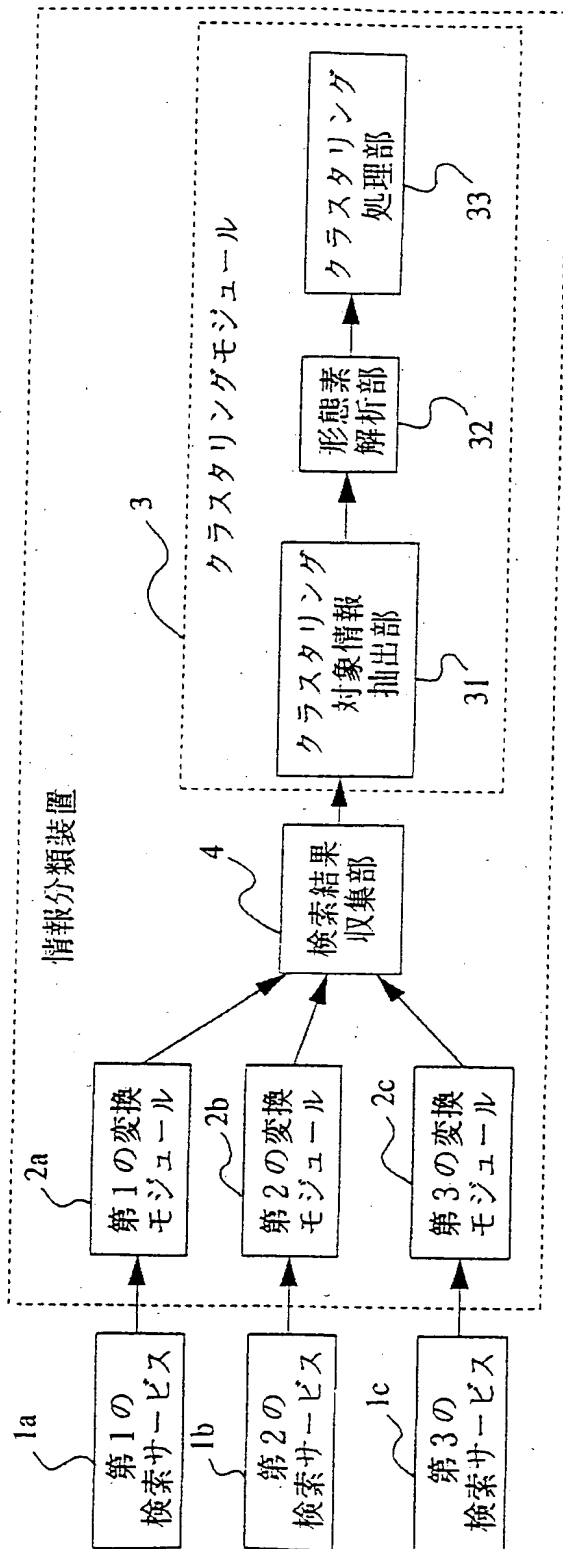


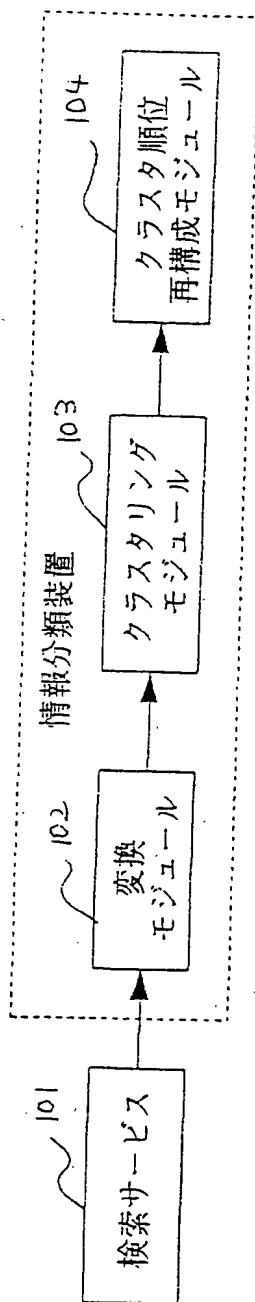
特徴要素	文書 D 1	文書 D 2	文書 D 3	文書 D 4	文書 D 5	文書 D 6	文書 D 7
用紙	1			1		1	1
カセット	1			1			1
増設		1	1		1		1

特徴要素	クラスタ
用紙	D1, D4, D6, D7
カセット	D1, D4, D7
増設	D2, D3, D5, D7

特徴要素	タイトル
用紙	<ul style="list-style-type: none"> * 用紙カセットについて * 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて * 印刷後における出力された用紙の汚れについて * 用紙カセットの増設について
カセット	<ul style="list-style-type: none"> * 用紙カセットについて * 用紙設定で「トレイ」「カセット」「ジドウ」の切替えについて * 用紙カセットの増設について
増設 ～	<ul style="list-style-type: none"> * レーザープリンタのメモリ増設について * オプティカルインターフェースカードの増設について * プリンタにフォントを追加するためにハードディスクを増設する場合について * 用紙カセットの増設について







クラス名		文書数	スコア	文書のタイトル
				検索結果
機型		16		133 東北エプソンの概要 49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 33 セイコーエプソン会社概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 (クラス内合計 579 平均 36)
LP		16		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 7 ニスバーレーザ LP-8300 (クラス内合計 450 平均 28)
仕様		14		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 (クラス内合計 413 平均 29)
デバイス		9		117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP (クラス内合計 849 平均 94)
半導体		7		157 半導体半導体環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP (クラス内合計 757 平均 108)
電子		4		111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ (クラス内合計 349 平均 87)

図 1 2

12/17

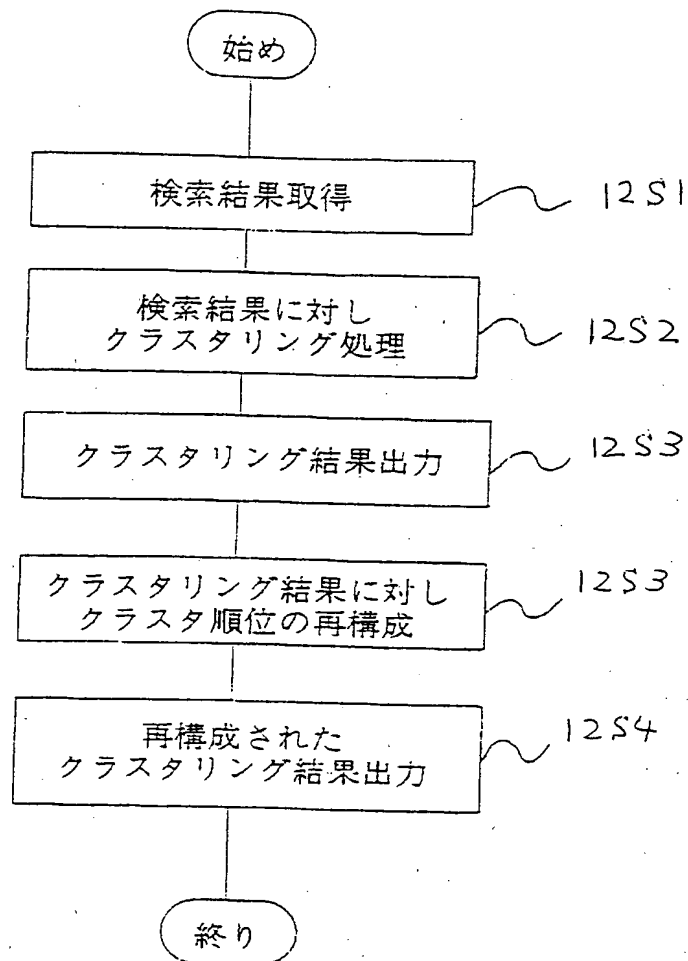


図 1 3

スコア		文書のタイトル
クラス名	文書数	検本結果
半導体	7	157 半導体事業部環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP (クラス内合計 757 平均 108)
デバイス	9	117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP (クラス内合計 849 平均 94)
電子	4	111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ (クラス内合計 349 平均 87)
概要	16	133 東北ニブソンの概要 49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 33 セイコーニブソン会社概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 (クラス内合計 579 平均 36)
仕様	14	49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 (クラス内合計 413 平均 29)
LP	16	49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-8300PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 7 エンバーレーザ LP-8300 (クラス内合計 450 平均 28)

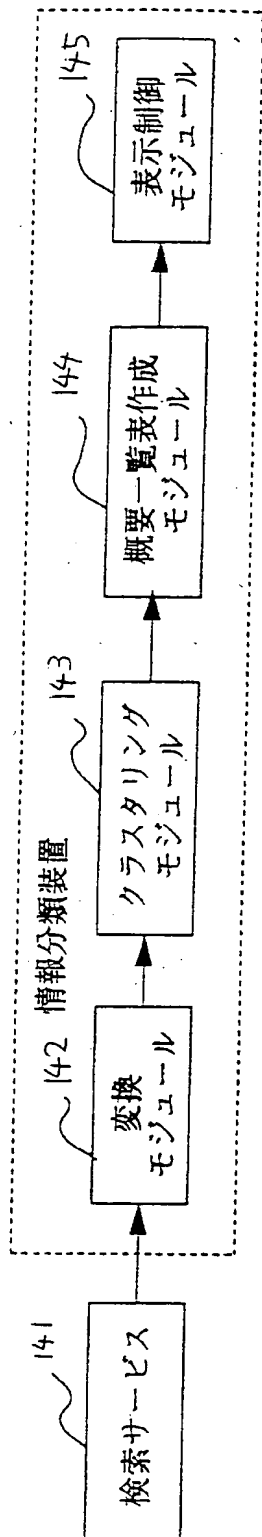


図 1 5

15/17

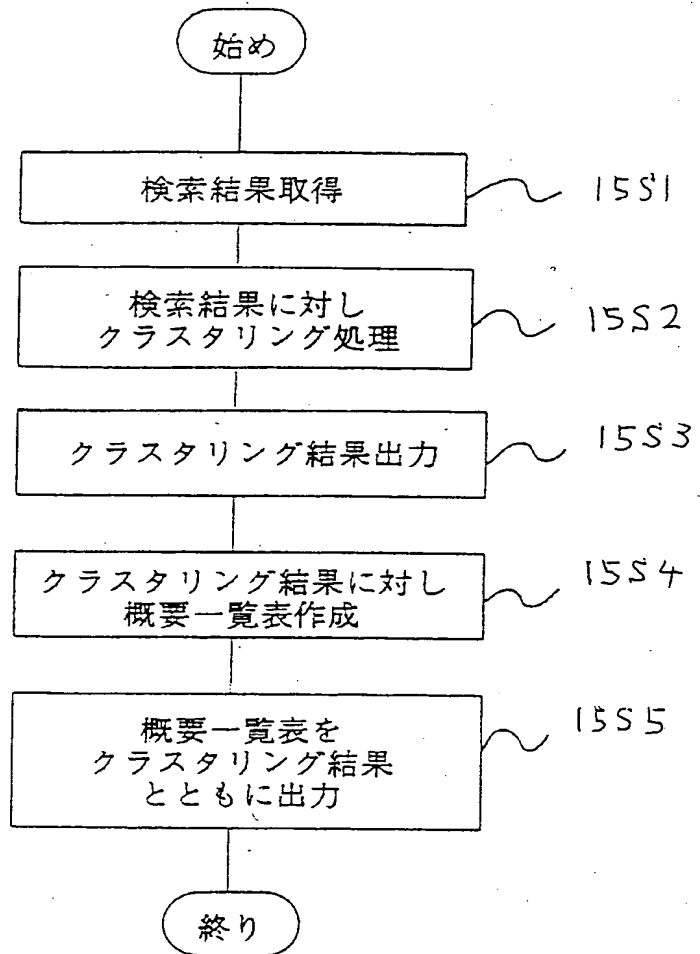


図 16

16/17

クラスタリング結果概要一覧					
概要	LP	仕様	デバイス	半導体	電子

1610

1620

クラス名		文書数	スコア	文書のタイトル
概要		16		133 東北ニアソンの概要 49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 33 セイコーニアソン会社概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-830PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要
LP		16		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-700S Spec Sheet & Option 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-830PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要 7 ニスバーレーダー LP-8300
仕様		14		49 LP-7000 仕様概要 38 LP-8000 仕様概要 38 LP-8400 仕様概要 34 LP-700 仕様概要 30 LP-9200 仕様概要 28 LP-9200PS2 仕様概要 28 LP-830PS 仕様概要 27 LP-8300 仕様概要 26 LP-8600 仕様概要 24 LP-8200 仕様概要 23 LP-710 仕様概要 23 LP-800 仕様概要 23 LP-900 仕様概要 22 LP-500 仕様概要
デバイス		9		117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 101 デバイス コラボレーション 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ 56 デバイス 半導体 ASSP
半導体		7		157 半導体事業部環境方針 117 デバイス 半導体 CARD-PC 115 デバイス 半導体 メモリ 111 電子デバイス ASIC 111 デバイス 半導体 PCカード製品 90 電子デバイス 半導体 マイコン 56 デバイス 半導体 ASSP
電子		4		111 電子デバイス ASIC 90 電子デバイス 半導体 マイコン 82 EPSON 電子デバイス新製品 66 EPSON 電子デバイスお問い合わせ

図 17

17/17

クラスタリング結果一覧		
www.epson.co.jp	www.i-love-epson.co.jp	その他

1740

1730

クラス名	文書数	文書のタイトル	検索結果
www.epson.co.jp	31		<p>半導体事業部 環境方針</p> <p>先登からのメッセージ-テキスト</p> <p>http://www.epson.co.jp/epson/ac/saiyou/career/</p> <p>http://www.epson.co.jp/epson/ac/saiyou/career/</p> <p>東北エプソンの概要</p> <p>「vividly NOTE」</p> <p>http://www.epson.co.jp/epson/ac/saiyou/career/</p> <p>東北エプソン 先登からのメッセージ</p> <p>Lamtec社とセイコーエプソン、半導体ビジョ</p> <p>東北エプソン 企業データ</p> <p>デバイス-半導体-CARD-PC</p> <p>半導体所製環境方針</p> <p>デバイス-半導体-メモリ</p> <p>http://www.epson.co.jp/epson/ac/saiyou/new/h9/</p> <p>http://www.epson.co.jp/epson/ac/saiyou/new/h9/</p> <p>職内内容</p> <p>電子デバイス-半導体-ASIC</p> <p>デバイス-半導体-PCカード製品</p> <p>http://www.epson.co.jp/epson/japanese/seihin/de</p> <p>東北エプソン 基本ポリシー</p> <p>酒田 5 号棟竣工</p> <p>デバイス コラボレーション ページ</p> <p>EPSON ELECTRONIC DEVICES INDEX</p> <p>東北エプソン株式会社 環境方針</p> <p>ムッシュ</p> <p>トランに新会社Shanghai Epson Electronicsを設</p> <p>EPSON Ecology</p> <p>電子デバイス-半導体-マイクロコンピニ</p> <p>SEIKO EPSON GROUP</p> <p>http://www.epson.co.jp/epson/japanese/seihin/de</p> <p>EPSON 電子デバイス新製品</p>
www.i-love-epson.co.jp	14		<p>LP-7000 仕様概要</p> <p>LP-8000 仕様概要</p> <p>LP-8400 仕様概要</p> <p>LP-700 仕様概要</p> <p>LP-9200 仕様概要</p> <p>LP-9200PS2 仕様概要</p> <p>LP-830PS 仕様概要</p> <p>LP-8300 仕様概要</p> <p>LP-8600 仕様概要</p> <p>LP-8200 仕様概要</p> <p>LP-710 仕様概要</p> <p>LP-800 仕様概要</p> <p>LP-900 仕様概要</p> <p>LP-500 仕様概要</p>
その他			<p>口内環境問題：オゾン層破壊、オゾン層とは</p> <p>KX-Server: PP-430</p>

特許協力条約に基づく国際出願願書

F005275W000

原本（出願用） - 印刷日時 2000年06月01日（01.06.2000）木曜日 12時05分42秒

0	受理官庁記入欄	
0-1	国際出願番号.	
0-2	国際出願日	
0-3	(受付印)	
0-4	様式-PCT/RO/101 この特許協力条約に基づく国際出願願書は、 右記によって作成された。	PCT-EASY Version 2.90 (updated 08.03.2000)
0-5	申立て 出願人は、この国際出願が特許協力条約に従って処理されることを請求する。	
0-6	出願人によって指定された受理官庁	日本国特許庁 (RO/JP)
0-7	出願人又は代理人の書類記号	F005275W000
I	発明の名称	情報分類方法及び情報分類装置並びに情報分類処理プログラムを記録した記録媒体
II	出願人	出願人である (applicant only)
II-1	この欄に記載した者は	米国を除くすべての指定国 (all designated States except US)
II-2	右の指定国についての出願人である。	
II-4ja	名称	セイコーエプソン株式会社
II-4en	Name	SEIKO EPSON CORPORATION
II-5ja	あて名:	163-0811 日本国 東京都 新宿区 西新宿 2 丁目 4 番 1 号
II-5en	Address:	4-1, Nishi-Shinjuku 2-Chome Shinjuku-Ku, Tokyo 163-0811 Japan
II-6	国籍 (国名)	日本国 JP
II-7	住所 (国名)	日本国 JP
II-8	電話番号	03-3348-3114
II-9	ファクシミリ番号	03-3340-4258
III-1	その他の出願人又は発明者	出願人及び発明者である (applicant and inventor)
III-1-1	この欄に記載した者は	米国のみ (US only)
III-1-2	右の指定国についての出願人である。	
III-1-4ja	氏名(姓名)	長石 道博
III-1-4en	Name (LAST, First)	NAGAISHI, Michihiro
III-1-5ja	あて名:	392-8502 日本国 長野県 諏訪市 大和 3 丁目 3 番 5 号
III-1-5en	Address:	セイコーエプソン株式会社内 c/o SEIKO EPSON CORPORATION 3-5, Owa 3-Chome, Suwa-Shi, Nagano 392-8502 Japan
III-1-6	国籍 (国名)	日本国 JP
III-1-7	住所 (国名)	日本国 JP

III-2 III-2-1 III-2-2 III-2-4ja III-2-4en III-2-5ja	その他の出願人又は発明者 この欄に記載した者は 右の指定国についての出願人である。 氏名(姓名) Name (LAST, First) あて名:	出願人及び発明者である (applicant and inventor) 米国のみ (US only) 三輪 真司 MIWA, Shinji 392-8502 日本国 長野県 諏訪市 大和3丁目3番5号 セイコーエプソン株式会社内 c/o SEIKO EPSON CORPORATION 3-5, Owa 3-Chome, Suwa-shi, Nagano 392-8502 Japan
III-2-5en	Address:	
III-2-6 III-2-7	国籍(国名) 住所(国名)	日本国 JP 日本国 JP
IV-1 IV-1-1ja IV-1-1en IV-1-2ja	代理人又は共通の代表者、通知 のあて名 下記の者は国際機関において右 記のごとく出願人のために行動 する。 氏名(姓名) Name (LAST, First) あて名:	代理人 (agent) 鈴木 喜三郎 SUZUKI, Kisaburo 392-8502 日本国 長野県 諏訪市 大和3丁目3番5号 セイコーエプソン株式会社 知的財産部内 c/o Intellectual Property Department SEIKO EPSON CORPORATION 3-5, Owa 3-chome Suwa-Shi, Nagano 392-8502 Japan
IV-1-2en	Address:	
IV-1-3 IV-1-4	電話番号 ファクシミリ番号	0266-52-3139 0266-58-3243
IV-2 IV-2-1ja IV-2-1en	その他の代理人 氏名 Name(s)	筆頭代理人と同じあて名を有する代理人 (additional agent(s) with same address as first named agent) 上柳 雅誉; 須澤 修 KAMIYANAGI, Masataka; SUZAWA, Osamu
V V-1	国の指定 広域特許 (他の種類の保護又は取扱いを 求める場合には括弧内に記載す る。)	AP: GH GM KE LS MW SD SL SZ TZ UG ZW 及びハラレプロトコルと特許協力条約の締約国であ る他の国 EA: AM AZ BY KG KZ MD RU TJ TM 及びユーラシア特許条約と特許協力条約の締約国で ある他の国 EP: AT BE CH&LI CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE 及びヨーロッパ特許条約と特許協力条約の締約国であ る他の国 OA: BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG 及びアフリカ知的所有権機構と特許協力条約の締約国 である他の国

特許協力条約に基づく国際出願願書

F005275W000




原本（出願用） - 印刷日時 2000年06月01日（01.06.2000）木曜日 12時05分42秒

V-2	国内特許 (他の種類の保護又は取扱いを 求める場合には括弧内に記載す る。)	AE AG AL AM AT AU AZ BA BB BG BR BY CA CH&LI CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW	
V-5	指定の確認の宣言 出願人は、上記の指定に加えて 、規則4.9(b)の規定に基づき、 特許協力条約のもとで認められ る他の全ての国の指定を行う。 ただし、V-6欄に示した国の指 定を除く。出願人は、これらの 追加される指定が確認を条件と していること、並びに優先日から 15月が経過する前にその確認 がなされない指定は、この期間 の経過時に、出願人によって取 り下げられたものとみなされる ことを宣言する。		
V-6	指定の確認から除かれる国	なし (NONE)	
VI-1	先の国内出願に基づく優先権主 張		
VI-1-1	先の出願日	1999年06月04日 (04.06.1999)	
VI-1-2	先の出願番号	特願平11-158497	
VI-1-3	国名	日本国 JP	
VI-2	先の国内出願に基づく優先権主 張		
VI-2-1	先の出願日	1999年06月17日 (17.06.1999)	
VI-2-2	先の出願番号	特願平11-171723	
VI-2-3	国名	日本国 JP	
VI-3	先の国内出願に基づく優先権主 張		
VI-3-1	先の出願日	1999年07月07日 (07.07.1999)	
VI-3-2	先の出願番号	特願平11-193141	
VI-3-3	国名	日本国 JP	
VI-4	優先権証明書送付の請求 上記の先の出願のうち、右記の 番号のものについては、出願書 類の認証謄本を作成し国際事務 局へ送付することを、受理官庁 に対して請求している。	VI-1, VI-2, VI-3	
VII-1	特定された国際調査機関(ISA)	日本国特許庁 (ISA/JP)	
VIII	照合欄	用紙の枚数	添付された電子データ
VIII-1	願書	4	-
VIII-2	明細書	34	-
VIII-3	請求の範囲	5	-
VIII-4	要約	1	f005275wo00.txt
VIII-5	図面	17	-
VIII-7	合計	61	
VIII-8	添付書類	添付	添付された電子データ
VIII-8	手数料計算用紙	✓	-
VIII-9	別個の記名押印された委任状	✓	-
VIII-16	PCT-EASYディスク	-	フレキシブルディスク
VIII-17	その他	納付する手数料に相当す る特許印紙を貼付した書 面	-

特許協力条約に基づく国際出願願書

F005275W000

原本（出願用） - 印刷日時 2000年06月01日（01.06.2000）木曜日 12時05分42秒

VIII-18	要約書とともに提示する図の番号	1
VIII-19	国際出願の使用言語名:	日本語 (Japanese)
IX-1	提出者の記名押印	
IX-1-1	氏名(姓名)	鈴木 喜三郎
IX-2	提出者の記名押印	
IX-2-1	氏名(姓名)	上柳 雅誉
IX-3	提出者の記名押印	
IX-3-1	氏名(姓名)	須澤 修

受理官庁記入欄

10-1	国際出願として提出された書類の実際の受理の日	
10-2	図面:	
10-2-1	受理された	
10-2-2	不足図面がある	
10-3	国際出願として提出された書類を補完する書類又は図面であつてその後期間内に提出されたものの実際の受理の日(訂正日)	
10-4	特許協力条約第11条(2)に基づく必要な補完の期間内の受理の日	
10-5	出願人により特定された国際調査機関	ISA/JP
10-6	調査手数料未払いにつき、国際調査機関に調査用写しを送付していない	

国際事務局記入欄

11-1	記録原本の受理の日	
------	-----------	--

PATENT COOPERATION TREATY

PCT

NOTICE INFORMING THE APPLICANT OF THE
COMMUNICATION OF THE INTERNATIONAL
APPLICATION TO THE DESIGNATED OFFICES

(PCT Rule 47.1(c), first sentence)

From the INTERNATIONAL BUREAU

To:
SUZUKI, Kisaburo
Intellectual Property Department
Seiko Epson Corporation
3-5, Owa 3-chome
Suwa-shi, Nagano 392-8502
JAPON

RECEIVED

DEC 26 2000

Intellectual Property Dept

SEIKO EPSON

IMPORTANT NOTICE

Date of mailing (day/month/year) 14 December 2000 (14.12.00)			
Applicant's or agent's file reference F005275WO00			
International application No. PCT/JP00/03623	International filing date (day/month/year) 02 June 2000 (02.06.00)	Priority date (day/month/year) 04 June 1999 (04.06.99)	
Applicant SEIKO EPSON CORPORATION et al			

1. Notice is hereby given that the International Bureau has communicated, as provided in Article 20, the international application to the following designated Offices on the date indicated above as the date of mailing of this Notice:
AG,AU,DZ,KP,KR,US

In accordance with Rule 47.1(c), third sentence, those Offices will accept the present Notice as conclusive evidence that the communication of the international application has duly taken place on the date of mailing indicated above and no copy of the international application is required to be furnished by the applicant to the designated Office(s).

2. The following designated Offices have waived the requirement for such a communication at this time:
AE,AL,AM,AP,AT,AZ,BA,BB,BG,BR,BY,CA,CH,CN,CR,CU,CZ,DE,DK,DM,EA,EE,EP,ES,FI,GB,GD,
GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,
NO,NZ,OA,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,UZ,VN,YU,ZA,ZW
The communication will be made to those Offices only upon their request. Furthermore, those Offices do not require the applicant to furnish a copy of the international application (Rule 49.1(a-bis)).

3. Enclosed with this Notice is a copy of the international application as published by the International Bureau on
14 December 2000 (14.12.00) under No. WO 00/75809

REMINDER REGARDING CHAPTER II (Article 31(2)(a) and Rule 54.2)

If the applicant wishes to postpone entry into the national phase until 30 months (or later in some Offices) from the priority date, a demand for international preliminary examination must be filed with the competent International Preliminary Examining Authority before the expiration of 19 months from the priority date.

It is the applicant's sole responsibility to monitor the 19-month time limit.

Note that only an applicant who is a national or resident of a PCT Contracting State which is bound by Chapter II has the right to file a demand for international preliminary examination.

REMINDER REGARDING ENTRY INTO THE NATIONAL PHASE (Article 22 or 39(1))

If the applicant wishes to proceed with the international application in the national phase, he must, within 20 months or 30 months, or later in some Offices, perform the acts referred to therein before each designated or elected Office.

For further important information on the time limits and acts to be performed for entering the national phase, see the Annex to Form PCT/IB/301 (Notification of Receipt of Record Copy) and Volume II of the PCT Applicant's Guide.

The International Bureau of WIPO 34, chemin des Colombettes 1211 Geneva 20, Switzerland	Authorized officer J. Zahra
Facsimile No. (41-22) 740.14.35	Telephone No. (41-22) 338.83.38

PCT REQUEST

1/5

F005275W000

0	For receiving Office use only	
0-1	International Application No.	
0-2	International Filing Date	
0-3	Name of receiving Office and "PCT International Application"	
0-4	Form - PCT/RO/101 PCT Request	
0-4-1	Prepared using	PCT-EASY Version 2.91 (updated 01.01.2001)
0-5	Petition The undersigned requests that the present international application be processed according to the Patent Cooperation Treaty	
0-6	Receiving Office (specified by the applicant)	Japanese Patent Office (RO/JP)
0-7	Applicant's or agent's file reference	F005275W000
I	Title of invention	INFORMATION SORTING METHOD, INFORMATION SORTING APPARATUS, AND STORAGE MEDIUM FOR STORING AN INFORMATION SORTING PROCESSING SOFTWARE PROGRAM
II	Applicant	
II-1	This person is:	applicant only
II-2	Applicant for	all designated States except US
II-4	Name	SEIKO EPSON CORPORATION
II-5	Address:	4-1, Nishi-Shinjuku 2-Chome Shinjuku-Ku, Tokyo 163-0811 Japan
II-6	State of nationality	JP
II-7	State of residence	JP
II-8	Telephone No.	03-3348-3114
II-9	Facsimile No.	03-3340-4258
III-1	Applicant and/or inventor	
III-1-1	This person is:	applicant and inventor
III-1-2	Applicant for	US only
III-1-4	Name (LAST, First)	NAGAISHI, Michihiro
III-1-5	Address:	c/o SEIKO EPSON CORPORATION 3-5, Owa 3-Chome Suwa-Shi, Nagano 392-8502 Japan
III-1-6	State of nationality	JP
III-1-7	State of residence	JP

III-2	Applicant and/or inventor	
III-2-1	This person is:	applicant and inventor
III-2-2	Applicant for	US only
III-2-4	Name (LAST, First)	MIWA, Shinji
III-2-5	Address:	c/o SEIKO EPSON CORPORATION 3-5, Owa 3-Chome Suwa-shi, Nagano 392-8502 Japan
III-2-6	State of nationality	JP
III-2-7	State of residence	JP
IV-1	Agent or common representative; or address for correspondence The person identified below is hereby/has been appointed to act on behalf of the applicant(s) before the competent International Authorities as:	agent
IV-1-1	Name (LAST, First)	SUZUKI, Kisaburo
IV-1-2	Address:	c/o Intellectual Property Department SEIKO EPSON CORPORATION 3-5, Owa 3-Chome Suwa-Shi, Nagano 392-8502 Japan
IV-1-3	Telephone No.	0266-52-3139
IV-1-4	Facsimile No.	0266-58-3243
IV-2	Additional agent(s)	additional agent(s) with same address as first named agent
IV-2-1	Name(s)	KAMIYANAGI, Masataka; SUZAWA, Osamu
V	Designation of States	
V-1	Regional Patent (other kinds of protection or treatment, if any, are specified between parentheses after the designation(s) concerned)	AP: GH GM KE LS MW SD SL SZ TZ UG ZW and any other State which is a Contracting State of the Harare Protocol and of the PCT (except MZ) EA: AM AZ BY KG KZ MD RU TJ TM and any other State which is a Contracting State of the Eurasian Patent Convention and of the PCT EP: AT BE CH&LI CY DE DK ES FI FR GB GR IE IT LU MC NL PT SE and any other State which is a Contracting State of the European Patent Convention and of the PCT (except TR) OA: BF BJ CF CG CI CM GA GN GW ML MR NE SN TD TG and any other State which is a member State of OAPI and a Contracting State of the PCT

V-2	National Patent (other kinds of protection or treatment, if any, are specified between parentheses after the designation(s) concerned)	AE AG AL AM AT AU AZ BA BB BG BR BY CA CH&LI CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MA MD MG MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW
V-5	Precautionary Designation Statement In addition to the designations made under items V-1, V-2 and V-3, the applicant also makes under Rule 4.9(b) all designations which would be permitted under the PCT except any designation(s) of the State(s) indicated under item V-6 below. The applicant declares that those additional designations are subject to confirmation and that any designation which is not confirmed before the expiration of 15 months from the priority date is to be regarded as withdrawn by the applicant at the expiration of that time limit.	
V-6	Exclusion(s) from precautionary designations	NONE
VI-1	Priority claim of earlier national application	
VI-1-1	Filing date	04 June 1999 (04.06.1999)
VI-1-2	Number	11-158497(P)
VI-1-3	Country	JP
VI-2	Priority claim of earlier national application	
VI-2-1	Filing date	17 June 1999 (17.06.1999)
VI-2-2	Number	11-171723(P)
VI-2-3	Country	JP
VI-3	Priority claim of earlier national application	
VI-3-1	Filing date	07 July 1999 (07.07.1999)
VI-3-2	Number	11-193141(P)
VI-3-3	Country	JP
VI-4	Priority document request The receiving Office is requested to prepare and transmit to the International Bureau a certified copy of the earlier application(s) identified above as item(s):	VI-1, VI-2, VI-3
VII-1	International Searching Authority Chosen	Japanese Patent Office (JPO) (ISA/JP)

VIII	Check list	number of sheets	electronic file(s) attached
VIII-1	Request	4	-
VIII-2	Description	34	-
VIII-3	Claims	5	-
VIII-4	Abstract	1	f005275wo00.txt
VIII-5	Drawings	17	-
VIII-7	TOTAL	61	
	Accompanying items	paper document(s) attached	electronic file(s) attached
VIII-8	Fee calculation sheet	✓	-
VIII-9	Separate signed power of attorney	✓	-
VIII-16	PCT-EASY diskette	-	diskette
VIII-17	Other (specified):	Revenue stamps of transmittal fee for receiving office	-
VIII-18	Figure of the drawings which should accompany the abstract	1	
VIII-19	Language of filing of the international application	Japanese	
IX-1	Signature of applicant or agent		
IX-1-1	Name (LAST, First)	SUZUKI, Kisaburo	
IX-2	Signature of applicant or agent		
IX-2-1	Name (LAST, First)	KAMIYANAGI, Masataka	
IX-3	Signature of applicant or agent		
IX-3-1	Name (LAST, First)	SUZAWA, Osamu	

FOR RECEIVING OFFICE USE ONLY

10-1	Date of actual receipt of the purported international application	
10-2	Drawings:	
10-2-1	Received	
10-2-2	Not received	
10-3	Corrected date of actual receipt due to later but timely received papers or drawings completing the purported international application	
10-4	Date of timely receipt of the required corrections under PCT Article 11(2)	
10-5	International Searching Authority	ISA/JP
10-6	Transmittal of search copy delayed until search fee is paid	

PCT REQUEST

5/5

F005275WO00

FOR INTERNATIONAL BUREAU USE ONLY

11-1	Date of receipt of the record copy by the International Bureau	
------	---	--